



The Open Society & Its Complexities

Gerald Gaus

The Open Society and Its Complexities

PHILOSOPHY, POLITICS, AND ECONOMICS

Ryan Muldoon, Carmen Pavel,
Geoff Sayre-McCord, Eric Schliesser, Itai Sher

Series Editors

Published in the Series

The Open Society and Its Complexities
Gerald Gaus

The Open Society and Its Complexities

GERALD GAUS

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2021

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Names: Gaus, Gerald, author.

Title: The open society and its complexities / Gerald Gaus.

Description: New York, N.Y. : Oxford University Press, 2021. |

Series: Philosophy, Politics, and Economics | Includes bibliographical references.

Identifiers: LCCN 2020058625 (print) | LCCN 2020058626 (ebook) |

ISBN 9780190648978 (hardback) | ISBN 9780190648992 (epub)

Subjects: LCSH: Social sciences—Philosophy. | Social ethics. | Liberalism. | Hayek, Friedrich A. von (Friedrich August), 1899–1992.

Classification: LCC H61 .G38 2021 (print) | LCC H61 (ebook) |

DDC 300.1—dc23

LC record available at <https://lcn.loc.gov/2020058625>

LC ebook record available at <https://lcn.loc.gov/2020058626>

DOI: 10.1093/oso/9780190648978.001.0001

1 3 5 7 9 8 6 4 2

Printed by Integrated Books International, United States of America

Contents

<i>Foreword</i>	vii
<i>Preface</i>	ix

PROLEGOMENON: HAYEK'S THREE UNSETTLING THESES

§1. Beyond Human Nature?	3
§2. Beyond Moral Justification?	6
§3. Beyond Human Governance?	10
§4. Three Inquiries on the Open Society	14

PART I. THE RISE OF A NORMATIVE SPECIES: ARE WE FIT FOR THE OPEN SOCIETY?

§5. A Natural History of Moral Order	19
§6. The “Starting Point”	20
§7. The Egalitarian Revolution	29
§8. Self-Interest, Reciprocity, and Altruism	37
§9. Internalized, Enforced, Social Rules	43
§10. The Other Side of Morality	53
§11. Cultural Evolution	63
§12. The Rise and (Partial) Fall of Inequality	75
§13. A Complex Moral Species	83

PART II. DIVERSITY AND SELF-ORGANIZED COMPLEXITY: IS THE OPEN SOCIETY BEYOND JUSTIFICATION?

§14. Liberalism and the Open Society	95
§15. Understanding Diversity	103
§16. Autocatalytic Diversity	109

§17. Diversity and Complexity	118
§18. The Morality of Self-Organization	128
§19. The Social Contract	139
§20. A Self-Organization Model	147
§21. Moral Diversity in the Open Society	159

PART III. THE COMPLEXITIES OF
SELF-GOVERNANCE: CAN THE OPEN SOCIETY
BE GOVERNED?

§22. Self-Governance	171
§23. Macro Control	184
§24. Macro Structure	194
§25. Strategic Dilemmas and Polycentricity	198
§26. Meso-Level Goal Pursuit	204
§27. Sectoral Policy	207
§28. Self-Governance from the Bottom Up: Simplifying the Problems of Governance	228
§29. Our Moral Nature and Governance in the Open Society	237
§30. Liberal Democracy	241
<i>Epilogue</i>	247
<i>Appendix A: Agent Types</i>	251
<i>Appendix B: Equations in Simulation §27</i>	255
<i>Bibliography</i>	257
<i>Index</i>	281

Foreword

Philosophy, Politics, and Economics (PPE) is coming of age as a distinct area of research of high social relevance. Modern societies face complex challenges, including those associated with inequality, trade, discrimination, disease, environmental degradation, immigration, pluralism, and technological change. PPE can foster a more robust interdisciplinary understanding of these challenges.

The Oxford University Press book series *Philosophy, Politics, and Economics* provides a home for PPE research ranging from fundamental theoretical issues to pressing questions of public policy. Across the board, the series aims to publish work that integrates philosophy, political science, and economics (and, indeed, all the social sciences) to tackle or solve important theoretical and policy questions. We are self-consciously ecumenical about the scope and focus of work in the series: PPE can and should range broadly and we are excited to explore its frontiers.

It is fitting that the inaugural book in the *Philosophy, Politics, and Economics* series is by Jerry Gaus, who was a crucial figure in developing PPE in North America. Jerry founded the journal *Politics, Philosophy & Economics*, he wrote the introductory text *On Philosophy, Politics, and Economics*, and he was instrumental in the PPE program at the University of Arizona, as well as being the first head of department for the Political Economy and Moral Science department. Jerry's work was groundbreaking in no small part because of its serious integration of ideas from philosophy and the social sciences to more adequately address complex social challenges. This final book from Jerry demonstrates the promise of PPE as a field.

Ryan Muldoon, Carmen Pavel, Geoff Sayre-McCord,
Eric Schliesser, and Itai Sher
Series Editors

Preface

The Open Society, Hayek thought, is a new form of civilization—one that in many ways has outstripped our moral and cognitive capacities. As I show here, Hayek's analyses were far ahead of the social theory of his day—and in many ways, of ours too. He advanced strikingly original and sophisticated theories of cultural and moral evolution, as well as social complexity and self-organization. This book is an extended engagement with his revolutionary, and seldom appreciated, research project. However, it is not a book about Hayek: my aim is not explication, defense, or critique. As a living research project, it ultimately does not much matter what Hayek said, but where the best evidence and analyses lead. No matter how prescient, Hayek's core work was done over a half a century ago. We have learned a tremendous amount in the last twenty years: my aim here is to consider Hayek's core theses in light of this exciting contemporary work—and my own analysis of social morality.

This project requires that I draw not simply on political philosophy, but on evolutionary analyses, primatology, anthropology, moral psychology, analyses of complex systems, experimental economics, studies of norms, economic development, policy studies, analyses of governance and collective action, randomized control trials, and much more. The argument ranges from more traditional social and political philosophy to reports on fieldwork and experiments through empirical evidence to formal modeling. I do not see this method as innovative—it's the way I was taught to do political theory by John Chapman. In his graduate seminars at the University of Pittsburgh, John would hand us a pile of five or so books, with the assignment to write a paper on them for next week to present to the seminar. Nothing was off the agenda. Not just Rawls, Nozick, Hayek, Oakeshott, Barry, Walzer, Shklar, Lukács, and Macpherson, but also North, Geertz, E. O. Wilson, Lindblom, Phelps Brown, Meade, Weber, Tiger and Fox, McNeil, Wiles, Demsetz, Janes . . . all—and piles more—were relevant to thinking about social and political life.

John had no moniker for this approach except “what the job is.” He did sometimes contrast it to a type of analytical philosophy that he dubbed “watchmaking”—creating a tiny, beautiful, pristine mechanism but knowing nothing about the world in which it operated. At Arizona we have called this more comprehensive approach “political economy and moral science”—the name of our new department devoted to interdisciplinary investigation of moral, social, and political life. The emerging discipline of “PPE” is another manifestation of this view about what the job is. This long-standing methodology—going back at least to Hume and Smith—stands in stark opposition to so much of what now falls under the rubric of “political philosophy” in philosophy departments, which has degenerated into reports of the intuitions and judgments of a section of upper-middle-class Americans and Europeans. There is clear evidence that this is a remarkably biased group, both in political beliefs¹ and more generally across

¹ Peters et al., “Ideological Diversity, Hostility, and Discrimination in Philosophy.”

a range of intuitions.² Manifestly parochial and biased intuitions are increasingly presented as compelling evidence of objective moral demands to which all must conform. It is not impossible that this distinctive subset of upper-middle-class Americans and Europeans have intuitions that provide privileged access into timeless justice, but I wouldn't bet the farm on it. If they are to be worthwhile and productive activities, social and political philosophy must carefully study humans and the immense diversity of their social lives before daring to reach normative conclusions. Of course, no matter what we do, our conclusions may fall into error and bias, but the more we see ourselves as students of society, rather than moral oracles and critics, the more likely it is that we will contribute to human knowledge and wisdom rather than use our professional status as a tool in ideological struggle.

Keynes famously remarked that “in the field of economic and political philosophy there are not many who are influenced by new theories after they are twenty-five or thirty years of age.”³ Whether political philosophy escapes its present arid self-righteousness depends on current graduate students, and whether enough find the world of inquiry sufficiently exciting to let go of reassuring certainties. To even begin to understand and evaluate the complex society we have jointly created, a wide diversity of approaches, tools, models, and perspectives are needed. Interdisciplinary and transdisciplinary work will be the norm, and in many respects research teams will replace the solo philosopher. These are the attractions of the resurgent moral science and PPE movements. As Keynes saw, their fate will be determined by graduate students.

I have been fortunate to have worked with a number of superlative graduate students who have, in their distinctive and diverse ways, appreciated the beauty of open inquiry, and pursued imaginative research agendas, usually taking them to places they never dreamed of. Jonathan Anomaly, Jacob Barrett, Piper Bringham, Adam Gjesdal, Alex Motchoulski, Keith Hankins, Brian Kogelmann, John Thrasher, Alex Schaeffer, Chad Van Schoelandt, and Kevin Vallier have all contributed to the evolution of this book over the last two decades. Their work has inspired me, and I have always learned from it. A great deal of the thinking for this book took place in graduate seminars, especially my seminar on moral and social evolution, and the seminar I co-taught with the amazing Jenann Ismael on diversity and complexity. Because the work on this book spanned two decades, I cannot even begin to thank all the people and audiences whose skeptical questioning constantly spurred me to rethink matters, and delve more deeply into the literature. Although Ken Binmore believes they are “tiresomely pedantic”⁴ (as I probably am), I have tried to carefully cite the people and work from whom I have learned, both to acknowledge how much I rely on them and to allow the reader to check my interpretations. Throughout the entire time, Fred D'Agostino often has been two steps ahead of me and was kind enough to look back to lend me a hand. Over the last decade I have learned heaps from two extraordinary and innovative philosophers, Ryan Muldoon and Shaun Nichols—especially over excellent IPAs. Paul Lewis's work on the development of Hayek's thinking on complexity

² Henrich, Heine, and Norenzayan, “The Weirdest People in the World?”

³ Keynes, *The General Theory*, pp. 383–4.

⁴ Binmore, *Crooked Thinking or Straight Talk?*, p. vii (reflexively confirming Binmore's point).

is so wonderful and thorough I largely abandoned my original exegetical aims, deferring to his far superior work. I received immensely valuable and challenging comments on the penultimate draft from Jonny Anomaly, Allen Buchanan, Jacob Barrett, Fred D'Agostino, and Shaun Nichols—my deep thanks to them, as well as to Oxford's earlier set of readers.

Open inquiry is both exiting and disconcerting: exciting because one can see that the view is being improved through the work of others, disconcerting because one finds claims that formerly seemed so right now cannot be supported. One must be prepared to dustbin cherished conclusions.⁵ Unlike political philosophies that pride themselves on being inoculated against facts, this means not only that further investigation may show that I am wrong about important matters—it means that it definitely will, and this inquiry would be a failure if it didn't. But it also means that, unlike so much contemporary current political philosophy, the analysis is open to constant improvement through appeal to improved models and better evidence. Developing a social and political philosophy truly appropriate to the complexities of the Open Society has barely begun: I intend this book as no more than a preliminary effort, to be carried on by those who know more.

⁵ To repeat an old joke—a Dean visiting a Math Department remarks, “What I like about your department is how efficient mathematicians are compared to scientists: give ‘em paper and a wastebasket and they’re all set. The only department that is more efficient is Philosophy—they don’t need the wastebasket.”

PROLEGOMENON
HAYEK'S THREE
UNSETTLING THESES

§1 Beyond Human Nature?

A mere two decades ago it was widely assumed that liberal democracy and the Open Society had decisively won their century-long struggle against authoritarianism and totalitarianism.¹ Alas, we now know better. In the United States, one political party seems captured by a nationalist populism, sometimes isolationist and at other times aggressive. Much of the “progressive” wing of the other is obsessed by the conviction that liberal capitalism is sexist, racist, oppressive, exploitative of the poor, and altogether unjust, despite the fact that in the last thirty years over a billion people have been raised from the miseries of extreme poverty,² while gender equality and inclusive diversity are at their apex in human history. Throughout Europe—from Scandinavia to Italy, from England to Poland—populism, authoritarianism, and hostility to diversity have become not only fixtures of political life, but in some countries have won government. Some liberal intellectuals are appalled, ready to blame the ignorance or viciousness of *hoi polloi*, while others have themselves become critics of liberalism and democracy. Meanwhile, liberation and democratization in China have reversed, and Russia is once again under the sway of authoritarianism.

Neither Hayek nor Popper would have been as deeply shocked by these developments as many of us who were complacent in thinking that the postwar victories of the Open Society were an irreversible achievement. Popper began *The Open Society and Its Enemies* by proclaiming that his aim was to wrestle with

difficulties faced by a civilization which aims at humaneness and reasonableness, at equality and freedom; a civilization which is still in its infancy, and which continues to grow in spite of the fact that it has been betrayed by so many of the intellectual leaders of mankind. It attempts to show that this civilization has not yet fully recovered from the shock of its birth—the transition from the tribal or “closed society,” with its submission to magical forces, to the “open society” which sets free the critical powers of man. It attempts to show that the shock of this transition is one of the factors that have made possible the rise of those reactionary movements which have tried, and still try, to overthrow civilization and to return to tribalism.³

As a philosopher, Popper attacked what he saw as the betrayal by other philosophers—most notably Plato, Hegel, and Marx—of the Open Society. Philosophy and enlightenment should be enemies of the superstitions of the closed society, yet in Popper’s eyes this trio turned philosophy to their defense.

Hayek’s analysis was importantly different and, I think, ultimately more insightful. For Hayek, the Open Society (or what he more often called “the Great Society”) was a precarious achievement because in many ways it is at odds with our deepest moral sentiments.

¹ For my own statement of this complacent view, see my “Liberalism at the End of the Century.”

² World Bank, “Decline of Global Extreme Poverty Continues but Has Slowed.”

³ Popper, *The Open Society and Its Enemies*, vol. I, p. 1.

Our attraction to “tribalism” is not a remnant of a superstitious or magical past, but at the core of our biologically evolved nature.

The transition from the small band to the settled community and finally the Open Society and with it to civilization was due to men learning to obey the same abstract rules instead of being guided by innate instincts to pursue common perceived goals. The innate natural longings were appropriate to the condition of life of the small band during which man had developed the neural structure which is still characteristic of *Homo sapiens*. These innate structures built into man’s organization in the course of perhaps 50,000 generations were adapted to a wholly different life from that he has made for himself during the last 500, or for most of us only 100, generations or so. It would probably be more correct to equate these “natural” instincts with “animal” rather than with characteristically human or good instincts. Indeed, the general use of “natural” as a term of praise is becoming very misleading, because one of the main functions of the rules learned later was to restrain the innate or natural instincts in the manner required to make the Great Society possible.⁴

During our long evolutionary history in small-scale bands, Hayek argued, we developed moral sentiments focusing on a parochial egalitarianism based on a close identity of interests. As a result, the attractions of “tribalism” are, in an important sense, *moral*: it strikes us as an appropriate and morally satisfying way to live. In contrast, Hayek hypothesized that our large-scale Open Society was the result of much more recent “cultural evolution” that produced a highly effective and welfare-enhancing system based on fidelity to impersonal rules. Thus it might be said that we have moved from parochial egalitarianism to the Open Society’s inclusive impartiality. About this, I believe, Hayek was surely correct.

How did this change from the parochial to the inclusive come about? Famously, Hayek proposed it evolved through a form of “group selection.” “The rules of conduct have . . . evolved because the groups who practiced them were more successful and displaced others.”⁵

Although the existence and preservation of the order of actions of a group can be accounted for only from the rules of conduct which individuals obey, these rules of conduct have developed because the individuals have been living in groups whose structures have gradually changed. In other words, the properties of the individuals which are significant for the existence and preservation of the group, and through this also for the existence and preservation of the individuals themselves, have been shaped by the selection of those individuals from the individuals living in groups which at each stage of evolution of the group tended to act according to such rules as made the group more efficient.⁶

Hayek perceived an enduring conflict between the “egalitarian–tribal” sentiments that arose in the distant past—which we might say are characteristic of the small-scale “closed

⁴ F. A. Hayek, *The Political Order of a Free People*, p. 160.

⁵ Hayek, *Rules and Order*, p. 18; Hayek, *The Fatal Conceit*, p. 25.

⁶ Hayek, “Notes on the Evolution of Systems of Rules of Conduct,” p. 283.

society”⁷—and the moral rules and attitudes necessary for the extended order of the Open Society. “One of the main functions of the rules learned later was to restrain the innate or natural instincts in the manner required to make the Great Society possible.”⁸ Our biologically evolved moral nature and the Open Society are thus in constant conflict, as nicely illustrated by Keynes’s famous evaluation of capitalism: “I think that capitalism, wisely managed, can probably be made more efficient for attaining economic ends than any alternative system yet in sight, but that in itself it is in many ways extremely objectionable.”⁹ Capitalism, and its impersonal, competitive, open, and, in some ways, non-egalitarian society produces the goods, but is in many ways morally objectionable. The real enemy of the Open Society, Hayek seems to tell us, is our own evolved moral sentiments, which make us so susceptible to calls from its foes.

Hayek’s startling, and deeply pessimistic, thesis was largely ignored, I think, because it violated so many of the intellectual taboos of its time. Most glaringly, in work from the 1960s through the mid-1980s, he maintained that our deepest moral sentiments were biologically evolved at a time when the very idea of an evolutionary explanation of our moral nature was scandalous in the academy. While the debate between “nature and nurture” accounts of human nature was part of the intellectual landscape of the 1960s and 1970s,¹⁰ within social philosophy most were strong advocates of the pervasive effects of environmental factors and learning: the natural selection of morality itself was an extreme outlier position.¹¹ It will be recalled that in 1975 E. O. Wilson published *Sociobiology*, and in 1978 *On Human Nature*, in which he argued that natural selection shaped important human behaviors, including ethics: he was accused of being a fascist and a racist, and was shouted down at Harvard.¹² Applications of evolutionary thought to social philosophy were inevitably associated with the discredited “Social Darwinism” of the late nineteenth century and, sure enough, when in the 1980s readers began to take note of the evolutionary dimension of Hayek’s work, both libertarian and social democrat reviewers applied that moniker to him—despite his careful explanation of why it was inappropriate.¹³ It was not until the last decades of the twentieth century—or perhaps even the first of this century—that evolutionary explanations of morality became first respectable and then, quickly, common.

A second reason that the evolutionary side of Hayek’s work was largely ignored was his emphasis on group selection. Sewall Wright, an advocate of group selection, participated in Hayek’s evolution seminar at Chicago during the 1950s, and it seems likely this was a major influence on his thinking.¹⁴ By the late 1970s, however, genetic group selection was widely seen as, though mathematically possible, practically almost

⁷ See Popper, *The Open Society and Its Enemies*, vol. I, p. 1.

⁸ Hayek, *The Political Order of a Free People*, p. 160. See also *The Fatal Conceit*, pp. 13ff.

⁹ John Maynard Keynes, “The End of Laissez-Faire,” p. 294.

¹⁰ On the minority, “nature,” side of the debate, the work of Hayek’s Austrian friend Konrad Lorenz was highly influential. See his *On Aggression*, first published in 1963. See also Tiger and Fox, *The Imperial Animal* (1971).

¹¹ See Pinker, *The Blank Slate*.

¹² Haidt notes this in *The Righteous Mind*, pp. 31–32; some of us vividly recall the abuse heaped on him.

¹³ See Paul, “Liberalism, Unintended Orders and Evolutionism”; Miller, “The Fatalistic Conceit.” For Hayek’s rejection of Social Darwinism, see Hayek, *Rules and Order*, p. 23; *The Fatal Conceit*, pp. 23ff.

¹⁴ See Caldwell, *Hayek’s Challenge*, p. 299.

impossible.¹⁵ The common view has been that the only phenotype relevant to natural selection is the individual organism; to greatly oversimplify—trait *t* will be selected if and only if individuals with *t* have a fitness advantage over non-*t* individuals. Again to woefully oversimplify: on a group selection view, an individual organism with trait *t* that has a fitness *dis*advantage vis à vis non-*t* individuals within its group can nevertheless be selected for if *groups* with a predominance of *t*-trait individuals outcompete *groups* with a predominance of non-*t* individuals.¹⁶ Being prepared to sacrifice your life for the group puts you at a disadvantage within the group vis à vis those who hold back, but a group of those prepared to sacrifice will displace a group predominated by those who hold back. Importantly, though, Hayek's crucial appeal to group selection was part of his account of cultural, not genetic, evolution, and it has been in models of cultural evolution that some version of group selection has become firmly entrenched in the last twenty years.¹⁷

The era of knee-jerk hostility to evolutionary explanations of morality and group-level selection (especially of cultures) has, thankfully, passed. We are now well positioned to re-evaluate Hayek's first unsettling thesis: a large body of theory and data on these issues has now been accumulated. In Part I of this book, I review much of this work. My aim is neither to praise nor debunk Hayek's unsettling thesis, but to evaluate it: is the Open Society opposed to our evolved moral and social sentiments? While taking the evolutionary roots of human sociality (and Hayek's analysis of it) seriously, I ultimately will be skeptical of sweeping claims about our "tribal" nature—claims that are becoming increasingly familiar.¹⁸ We shall see that humans are social creatures of many layers, from the Machiavellian, to the kin-focused, clan-favoring, tribal, inter-tribal, ethnic, national, and universalistic. The task of evaluating whether we are fit for the Open Society will be a study of an ambivalent species, which has evolved in—and has evolved—a striking array of social milieus, from autonomy-obsessed egalitarian bands to the despotic and ruthlessly hierarchical.

§2 Beyond Moral Justification?

2.1. Moral Complexity

Hayek repeatedly refers to "the twin ideas of evolution and spontaneous order."¹ On his analysis the forces of cultural evolution, operating largely through a form of group selection, have produced an unplanned or self-organized order. A system is

¹⁵ Critical here was Williams's (1966) *Adaptation and Natural Selection*. For perhaps the most important work reviving group selection in moral and social theory, see Sober and Wilson, *Unto Others*.

¹⁶ This was Darwin's view. See *The Descent of Man*, p. 498. For a sophisticated analysis of the group selection debate, see Strerelny, "The Return of the Group." See §11.4 in this volume.

¹⁷ The work of Boyd and Richerson was critical, though they developed a sophisticated model of gene-cultural coevolution. For an accessible overview, see Richerson and Boyd, *Not by Genes Alone*.

¹⁸ See, e.g., Goldberg, *Suicide of the West*; Greene, *Moral Tribes*.

¹ Hayek, "Dr. Bernard Mandeville," p. 250; "Notes on the Evolution of Systems of Rules of Conduct," p. 289. See also Hayek, *Rules and Order*, p. 23.

self-organizing when its elements behave in such a way as to accommodate their behavior to each other so as to achieve orderly patterns of relatively dense interactions without any central direction or planning. For Hayek our self-organizing social system resulting from cultural evolution is, formally, a complex system.² Complex structures are composed when many heterogeneous elements (e.g., agents), acting according to a set of rules, produce orderly patterns.³

A fundamental philosophical confusion, Hayek insists, is to treat the complex system of the evolved Open Society as if it were a simple determinate system. This confusion, he charges, is at the root of “constructivist” ethical theories, such as the social contract.⁴ This is surprising. Of all the traditions in social and political philosophy, the social contract would appear the most hospitable to the diversity of the Open Society. As Hayek stressed, in a diverse society there is no accepted single system of ends or ordering of values to determine the overall goodness of social states.⁵ While many varieties of ethical theory assert the existence of such a single system, it would seem that it was precisely the social contract tradition that understood disagreement about ends and values to be a permanent feature of modern society, to be accommodated rather than suppressed.⁶ Nevertheless, in Hayek’s eyes the social contract remains an instance of central (moral) planning. A central “controller”—aka the philosopher—reflects on the range of diversity in her society and comes up with a plan—aka the principles or institutions of justice—that will accommodate the observed “reasonable” diversity in an optimal way. On Hayek’s analysis the social contract theory exemplifies three fundamental errors.

- (i) Most obviously, it manifests the “*synoptic delusion* . . . the fiction that all the relevant facts are known to some one mind, and that it is possible to construct from this knowledge of the particulars a desirable social order.”⁷ No one can know the diversity of moral views, beliefs, and ideals of diverse orders of tens or hundreds of millions. The synoptic delusion is often hidden by normalizing assumptions, restricting the “contract” to a specific range of “reasonable” admissible views. This has the effect of appearing to render the calculations of the central controller (the philosopher) tractable while nevertheless claiming applicability to a much wider population.⁸
- (ii) Contract theory also supposes that we can know the justification of our rules: the purposes they serve, and whether they would be included in the social contract. Most contracts go beyond abstract principles to include evaluation of a set of institutions or practices.⁹ The crux of the contract is that individuals rationally

² See, for example, his “Notes on the Evolution of Systems of Rules of Conduct” and “The Theory of Complex Phenomena.”

³ See Hayek, “Rules, Perception and Intelligibility,” esp. pp. 245–8.

⁴ See, e.g., *Rules and Order*, pp. 10–1.

⁵ See, e.g., *The Road to Serfdom*, esp. chap. 5.

⁶ For arguments to this effect for Hobbes, Locke, Rousseau, and Kant, see the essays in Part I of Turner and Gaus (eds.), *Public Reason in Political Philosophy*.

⁷ Hayek, *Rules and Order*, p. 14.

⁸ See my *Tyranny of the Ideal*, pp. 150ff.

⁹ See Rawls, *A Theory of Justice*, Part II. G. A. Cohen sharply distinguishes such “rules of regulation” from core principles of justice (*Rescuing Justice and Equality*, e.g., pp. 323–7). However, they are central to the social contract theories of Hobbes, Locke, Rousseau, and Rawls.

judge whether a set of rules, practices, or institutions are acceptable; for that, the contractors must know how they function—the ends they are intended to serve. But Hayek insists that because our rules are part of an evolved system, it is exceedingly difficult—usually impossible—to grasp their functioning or purposes. Because of this, he indicates that we must view the rules as “purpose-independent.” Indeed, it is not even clear that it makes sense to ask for a rule’s justification.¹⁰

- (iii) Hobbesian versions of the social contract are so appealing to the philosophical mind because reason has a starring role in the analysis. Although strategic rationality is a cause of conflict in the state of nature, reason also points the way to society: it shows that we can only attain society by renouncing unconstrained self-interested maximization and binding ourselves into a cooperative, rule-based, truly social order. Captivated by this idea, a long line of distinguished moral and political theorists, right up to the present, have developed sophisticated accounts of politics, morality, and justice grounded on instrumental rationality.¹¹

However, once we take an evolutionary view, and understand humans as not only the creators, but the products of an evolved complex order, we see how the social contract’s instrumentalist constructivism leads us astray.

Societies differ from simpler complex structures by the fact that their elements are themselves complex structures whose chance to persist depends on (or at least is improved by) their being part of the more comprehensive structure. We have to deal here with integration on at least two different levels, with, on the one hand, the more comprehensive order assisting the preservation of ordered structures on the lower level, and, on the other, the kind of order which on the lower level determines the regularities of individual conduct assisting the prospect of the survival of the individual only through its effect on the overall order of the society. This means that the individual with a particular structure and behaviour owes its existence in this form to a society of a particular structure, because only within such a society has it been advantageous to develop some of its peculiar characteristics, while the order of society in turn is a result of these regularities of conduct which the individuals have developed in society.¹²

In this critical passage, Hayek stresses the interaction between two levels of selection regarding complexity. Cultural evolution created not only highly complex systems of social interactions, but also the sort of individuals who can successfully operate in such a complex culture. This idea, we shall see, is fundamental to more recent analysis of gene-culture evolution (§11.1): the evolution of culture, often as an adaptive response to competition with other cultures, selects for individual dispositions that thrive in that culture. Our moral rules evolved as an adaptive solution to problems of

¹⁰ Hayek, *Rules and Order*, pp. 81ff; *The Fatal Conceit*, pp. 66ff.

¹¹ See, e.g., Buchanan, *The Limits of Liberty*; Gauthier, *Morals by Agreement*, esp. chap. 3; Vanderschraaf, *Strategic Justice*; Moehler, *Minimal Morality*.

¹² Hayek, “Notes on the Evolution of Rules of Conduct,” pp. 287–8.

human cooperative life, *but so has our reason itself*. It, too, is an adaptive response to the rise of rules. Because our reason is itself the outcome of evolution, and because human social life is based on social rules, Hayek thus dismisses the Hobbesian conception of rationality as thoroughly instrumental: “Man is as much a rule-following animal as a purpose-seeking one.”¹³ Indeed, he insists that “the brain is an organ enabling us to absorb, but not to design culture.”¹⁴ And because social rules are a cultural phenomenon, they are transmitted by example and teaching; man is thus “distinguished” by “the capacity to imitate and to pass on what he has learned.”¹⁵ Since humans learn rules through imitation and example rather than through rational reconstruction, we typically do not fully understand what the rules accomplish or why they are needed. “In other words, man has certainly more often learnt to do the right thing without comprehending why it is the right thing, and he is still often served better by custom than by understanding.”¹⁶

According to claim (i), then, no single mind can comprehend the diversity of a complex society; according to (ii) we seldom know the justification for our social rules, and according to (iii) devising and knowing justifications is typically unnecessary: we are as much rule-following conformists as rational justifiers. We are thus left in deep doubt: can the Open Society be justified, and can those who understand it coherently engage in moral criticism and reform?

2.2. Rethinking the Enterprise of Justifying the Open Society

Hayek was a pioneer in the theory of complex social systems;¹⁷ here too he raced too far ahead of the intellectual fashions of his day.¹⁸ Because so few readers grasped the importance of complexity,¹⁹ Hayek’s stress on the limits of reason and justification, and the concomitant importance of conformity to traditional rules, often led to reading him as a type of conservative. Liberals, after all, believe in reason and the

¹³ Hayek, *Rules and Order*, p. 11.

¹⁴ Hayek, *The Political Order of a Free People*, p. 157.

¹⁵ Ibid., pp. 156–7.

¹⁶ Ibid., p. 157.

¹⁷ In a recent series of wonderful essays, Paul Lewis has explored the development of Hayek’s theory of complex systems. See Lewis’s “Ontology and the History of Economic Thought,” “The Emergence of ‘Emergence’ in the Work of F. A. Hayek,” “Notions of Order and Process in Hayek,” “An Analytical Core for Sociology,” and “Emergent Properties in the Work of Friedrich Hayek.”

¹⁸ I believe he is still far ahead of most economists, whose commitment not simply to the usefulness of single equilibrium models, but to their fundamental truth, conveys an unshakable faith in a world of classical dynamics. For an alternative picture, see Alan Kirman and Rajiv Sethi, “Disequilibrium Adjustment and Economic Outcomes.”

¹⁹ This despite the fact that Hayek repeatedly discusses it throughout his work. For an important exception, see Karen Vaughn, “Hayek’s Theory of the Market Order as an Instance of the Theory of Complex, Adaptive Systems.” Remarkably, “complexity” does even show up in the index in some of the best books on Hayek’s liberalism, such as Kukathas’s *Hayek and Modern Liberalism* and Gray’s *Hayek on Liberty*. Shearmur’s book on the Hayekian research program contains only a few scattered references. See his *Hayek and After*, pp. 111, 166, 205, 237. On the similar feature of Boettke’s wonderful *F. A. Hayek*, see my “A Branch on the Mainline: Hayek’s Analysis of Complex Adaptive Systems.”

Enlightenment, so questioning them must show lack of fidelity to the liberal cause.²⁰ However, the development of theories of evolution and analyses of societies as complex systems has rapidly developed in the last twenty years and, again, we shall see that the sort of worries expressed by Hayek are a staple of the literature. This is not, of course, to say that his conclusions are accepted (though some are); but the challenges to traditional philosophic justifications posed by the complexity of rule systems among diverse agents, the importance of social conformity to complex cultural evolution, the limits of reasoned justification—all are basic to recent analyses of evolution and complexity. We can no longer simply dismiss Hayek's challenges to the possibility of reasoned justification of the Open Society as the doubts of a conservative or an idiosyncratic Austrian economist.

The challenges posed by Hayek's first and second theses are different. With the first, our concern is whether Hayek's evolutionary account is sound—that our evolved moral sentiments constantly cause us to rebel against the Open Society and resort to a “tribal” moral outlook. The second challenge is, having recognized that evolutionary processes have produced a complex culture and morality not fully intelligible to us, and that an evolved complex culture requires fidelity to such evolved norms, what type of justification is open to us? If—as I think is correct—even the traditional social contract analysis depends on implausible claims to knowledge and the powers of reason, what remains? The unsettling possibility raised by Hayek's analysis is that we live in an Open Society, but can say little about the justification of its rules except that they have evolved, and we have little solid grounds for improving them. Given that the Open Society seeks to free human intellect to search for new ways of living and thinking, it would be unsettling indeed if its nature was such that its members could say little in criticism or justification of its rules.

§3 Beyond Human Governance?

3.1. Complexity and Self-Governance

Social complexity arises because outcomes are dependent on multiple, interconnected variables:

The position will here frequently be that if we already knew the relevant laws, we could predict that if several hundred specified factors had the values $x_1, x_2, x_3, \dots, x_n$, then there would always occur $y_1, y_2, y_3, \dots, y_n$. But in fact all that our observation suggests may be that if x_1, x_2, x_3 , and x_4 , then there will occur either $(y_1 \text{ and } y_2)$ or $(y_1$

²⁰ This is essentially the view of Michael Freeden, who depicts Hayek as a “liberal pretender” whose understanding of order “bears an uncommonly close resemblance to a major conservative precept.” *Ideologies and Political Theory*, p. 311.

and y_3) or (y_2 and y_3), or some similar situation—perhaps that if x_1, x_2, x_3 , and x_4 , then there will occur some y_1 and y_2 between which either the relation P or the relation Q will exist. There may be no possibility of getting beyond this by means of observation, because it may in practice be impossible to test all the possible combinations of the factors $x_1, x_2, x_3, x_4, \dots x_n$. If in the face of the variety and complexity of such a situation our imagination cannot suggest more precise rules than those indicated, no systematic testing will help us over the difficulty.¹

Critical to Hayek's analysis of complex social systems, then, is our inherent inability to predict future specific states of the system and, so, to plan or control them with any degree of precision. They constantly throw up novel and unexpected results, and small disturbances can push the system along drastically different paths. Contrary to what is often thought, this does not lead Hayek to a generalized skepticism about social knowledge, but to an exploration of the type of science appropriate to complex systems, and the sort of knowledge such a science can provide.² Hayek constantly stresses that evolutionary theory is a far superior model to, say, physics, for an economics of complexity. Employing Newtonian physics, we can make relatively precise predictions of where a satellite will land; employing Darwinian theory, we cannot predict the next big thing in human evolution—as we cannot predict the next big thing in the economy.³

Now—and here we begin to see sharper differences between Popper's and Hayek's diagnoses of the threats to the Open Society—Hayek repeatedly stresses that an erroneous understanding of the social sciences is one of the major threats to the self-organizing complex order produced by evolutionary processes.⁴ Natural sciences such as classical physics are modeled on fairly simple, determinate systems. And because these sciences—especially the social scientist's understanding of them—involve determinate equations, in emulating them economics and other social sciences developed determinate accounts of the social order. General equilibrium models suppose that the economic system tends toward one outcome. Equipped with determinate models, economists and other social scientists offer specific policy advice; after all, they understand the social system as a determinate system, so theoretical knowledge of the system should provide leverage for control.

¹ Hayek, "Degrees of Explanation," p. 200.

² "It is, of course, a commonplace that as we move from mechanics through the other branches of physics through inorganic and organic chemistry to the 'more highly organized' systems of biology we move in an ascending order of essential complexity. It is perhaps less familiar, but I believe no less true, that the only criterion by which we in fact single out some phenomena as belonging to [classical] physics is that they are essentially simple, i.e., that their theoretical explanation can be achieved by models requiring only very few parts." Hayek, "Economists and Philosophers," p. 438. See also Hayek's Nobel Lecture, "The Pretense of Knowledge."

³ See Hayek, "The Theory of Complex Phenomena."

⁴ This is the main theme of Hayek's *The Counter-Revolution in Science*. We need to be careful: Popper is also critical of "scientism," and perceives less conflict between his and Hayek's views. See *The Open Society and Its Enemies*, vol. I, pp. 285–6. See also Popper's letter to Isaiah Berlin about the abuse of science, and where he agrees with Hayek, in Karl Popper, *After the Open Society*, pp. 199–204. My thanks to Piers Norris Turner for pointing out this letter.

For Popper, superstition's revolt against reason was the chief threat to the Open Society. For Hayek, it is the abuse of reason by scientists and social scientists (often in their guise as policy advisors) that poses the critical danger. Thus Hayek's third unsettling thesis: our idolization of the powers of reason and the conviction of the intelligentsia that they can apply the findings of social science to control our complex order are enemies of the Open Society.⁵ Just as our genetically evolved moral intuitions may be at odds with the life of the Open Society, our deep cognitive intuitions about what we can know, and the efficacy of our social plans, tempt us to treat a far-ranging highly complex order as if it were a simple determinative system, with levers that can be pulled by central controllers to secure predicted social results. But given measurement and computational limits, we cannot predict the path that a specific intervention may produce: "all that our observation suggests may be that if x_1 , x_2 , x_3 , and x_4 , then there will occur either (y_1 and y_2) or (y_1 and y_3) or (y_2 and y_3), or some similar situation—perhaps that if x_1 , x_2 , x_3 , and x_4 , then there will occur some y_1 and y_2 between which either the relation P or the relation Q will exist." In addition to these problems, if we seek to control our social complex system, we must measure the new system state and determine which changes were the result of the intervention, and which were the result of other endogenous and exogenous factors (and of course these typically interact). Because the system is produced by the tightly coupled interaction of its many parts, the controller cannot identify the causal effects of its intervention from a host of other exogenous and endogenous changes that have occurred between the intervention and the attempt to measure its outcome. The upshot is that the evaluation of interventions typically focuses on the most obvious and expected consequences—what is easily observed.⁶ Thus policymakers and advisors may be quite convinced that their interventions are subject to evaluation, but this is based on a radically incomplete picture of the new social state they have brought about.

The extensive, diverse Open Society forms a complex social system and thus we must ask: What might we mean by self-governance in a such a system? Perhaps, all we can accomplish is endless series of myopic interventions of varying degrees of success, solving some problems to at least some people's satisfaction while causing unknown other problems, with no one ever really having a firm basis for concluding that, overall, the social order has been improved. If so, it would seem that self-government, democratic or otherwise, is an atavistic illusion—a holdover from a simpler period in which we lived in societies that really could be controlled by a central authority. To be sure, synoptic control and myopic intervention by no means exhaust the options—we need to inquire into other modes of self-governance and determine whether they are consistent with the diversity and complexity of the social order of the Open Society.

Supposing we discover effective modes of self-governance in our complex Open Society, we still face the question of whether they are recognizably democratic. *Prima facie*, one might conclude that intelligent interventions in a complex system must be based on sophisticated models devised by scientists studying complex systems. Think

⁵ This, of course, was the crux of *The Road to Serfdom*.

⁶ See Mack, "What Is Seen and What Is Not Seen."

of weather and climate, which all investigators see as complex systems. As we know from models of climate change, large proportions of the democratic public can be highly skeptical of models that they cannot understand (and of which experts offer competing versions)—and if they do accept them, it is on the authority of experts. Thus it has been argued that our complex social order is, literally, too complex for the democratic public to intelligently govern, which for all practical purposes must be handed over to an expert bureaucracy.⁷

3.2 The Complexities of the Open Society

In different ways, Hayek's theses reiterate a single point: the diverse, complex, Open Society has evolved to the point where it has outstripped basic human inclinations and capacities. The first thesis argues that it is fundamentally at odds with our deepest moral intuitions; the second that it has outstripped our capacity to understand the function and justification of its constitutive rules; and the third that it has evolved beyond our governance. Because of the first, we are constantly tempted to morally renounce it and, on moralistic grounds, construct barriers to it. Because of the second, our attempts to reflect and reconstruct its rules lead to constant and unrelenting moral conflict; we seek to justify, and so present a series of justificatory schemes and criticisms of its basic moral structure. Yet we do not really understand how it works or what it does for us. And we seek to devise policies to improve its functioning, yet again we do not have the knowledge to competently do so, hence we are constantly disappointed by the last round of interventions and we blame the last government for its failures and broken promises. Perhaps, as seems to have happened today (as it has in the past), the people begin to distrust all claims to expertise and seek simpler, more intuitive, solutions.⁸

These symptoms are manifested in our daily politics. We are witnessing rebellions against diversity and openness, seemingly irresolvable disagreements about what would constitute a justified moral basis of our social life, and almost universal exasperation about the inability of our politics to produce the sorts of changes that groups seek and politicians have promised. In this prolegomenon I have argued that Hayek's diagnoses of these symptoms, long dismissed as idiosyncratic and obviously erroneous, resonate with a large body of contemporary scientific work. This is not to say they are correct—it is to say that they demand further investigation. Because Hayek built his analysis on evolutionary and complexity science, his analysis is inherently open to revision.

⁷ The idea that the elements of contemporary society are tightly interconnected in a complex system was employed as early as the 1930s as a justification for increased central planning. See Graham, *Toward a Planned Society*, p. 20. Cf. Hayek, *The Road to Serfdom*, chap. 5. Hayek (p. 91) quotes Mussolini's 1929 Grand Fascist Council Report: "We were the first to assert that the more complicated the forms assumed by civilization, the more restricted the freedom of the individual must become." Schumpeter argued that the liberal conception of democracy as rule by the will of the people was destined to be replaced by political competition to lead the bureaucratic, socialist, planning, state. *Capitalism, Socialism, and Democracy*, Part IV.

⁸ In supporting Brexit, Michael Gove declared, "people in this country have had enough of experts." *Financial Times*, June 13, 2016.

§4 Three Inquiries on the Open Society

This work is composed of three inquiries on the Open Society, each taking up one of Hayek's unsettling theses. Later inquiries build on the results of the earlier. The first inquiry (Part I) considers the core problem of our evolved moral nature, and whether we are by nature "tribal" and parochial egalitarians who rebel at diversity, inclusion, and impartiality. I shall present a great deal of varied evidence that, in my view, shows us to be an ambivalent species, both self-interested and cooperative, autonomous and controlling, torn between reconciliation and conflict, egalitarian and domineering, expanding the sphere of moral cooperation outward, though often drawing back to more intense local networks. None of this, I shall argue, shows that we are unfit for the Open Society, though we are certainly not optimized for it. In this first Part we will see that many of Hayek's critical evolutionary insights are supported by current research, but he was, in my view, insufficiently aware of the way in which the Open Society builds upon our evolved nature as reciprocating, norm-following cooperators and how it too expresses our deep egalitarianism.

The second inquiry (Part II) turns from a broad evolutionary history of egalitarian cooperation based on reciprocity to an analysis of the Open Society itself. It begins by contrasting the idea of the Open Society to a traditional, Millian liberalism, based on free inquiry but, we shall see, ultimately limited diversity. The Open Society, I shall argue, is a unique form of civilization characterized by a process of autocatalytic diversity: an ever-increasing range of diversity, propelling yet more diversity. Hayek was correct that, once it has reached a certain level of complexity, such a social system cannot be planned or even fully understood: there is no way to accurately predict which, of the indefinite possibilities, will be realized. However, I shall dispute Hayek's claim that the cooperative order of this ever-expanding diverse society is provided by rules that have been selected through competition with other groups. Instead, I shall argue that the order of the Open Society is produced by a form of self-organization, growing out of our evolved cooperative and egalitarian natures. This second inquiry concludes with a formal model of how the moral rules that structure the self-organization of the Open Society can themselves be self-organized, and in so doing secure public justification. Thus I respond to Hayek's worry that the Open Society is beyond justification, while accepting his critique of traditional forms of moral justification such as the social contract.

Part II stresses the advantages of a diverse and Open Society, but I cannot pretend to demonstrate that these advantages outweigh the costs in everyone's eyes. There will be no attempt to provide a definite defense of the deep diversity of the Open Society over the normalized diversity of Millian liberalism. Thus one aspect of Hayek's skeptical account of justification is certainly correct: we cannot provide a definitive justification of the Open Society *as such* over many conceivable alternatives. But the demand for justification of social order *sub specie aeternitatis* is, as Hayek so often insisted, itself an obsession of a rationalist, as if we can somehow choose the nature of our civilization. The philosophical project of deciding what type of society to choose from a list

of distinct and relatively comprehensive alternatives is absurd. We cannot choose the fundamental nature of our civilization or social order; and because a morality has a necessary functional role in social life, we cannot simply choose our morality from a menu. This is the insight behind Ken Binmore's audacious claim that "orthodox moral philosophy has gotten us nowhere because it asks the wrong questions. If morality evolved along with the human race, asking how we ought to live makes as much sense as asking what animals ought to exist, or which language we ought to speak."¹

Normative political philosophy, I shall argue, is an analysis of what Stuart Kauffman calls the "adjacent possible,"² or what Rawls deemed the "neighborhood" of existing institutions.³ What sorts of rules and institutions can be justified in *our* society—how we can reform *our* social order? The complexity and diversity of the Open Society is, in my view, a parameter on political philosophy, not a variable to be altered.⁴ Our lives are deeply intertwined in a complex order of wide and deep diversity. We need to know what functional social morality can be justified *in it*. I shall neither bemoan flawed human nature for being too weak to follow "true" justice, nor devise an ideal to prop up our spirits that moral salvation remains within our horizon. Instead, I shall inquire how we can live without oppression and subjugation in a complex and deeply divided world. If this question can be answered, we have all the justification that we can obtain—and need. This, to be sure, still leaves us with the paradox of justification *in* (as opposed to, *of*) the Open Society: How can we justify rules if we do not understand their role in our society? Answering that will take a bit of work.

While Part II's focus is on the self-organization and public justification of the complex Open Society, the third inquiry (Part III) is devoted to Hayek's third unsettling thesis—that it is beyond human governance. This third Part disaggregates the idea of "self-governance" into three dimensions: goal pursuit, solving strategic problems, and providing the framework for self-organization. It then considers the prospects for each of these at the macro (system) level, the meso (middle) level, and the micro level. My aim here is to advance beyond overly broad claims such that self-governance in complex systems is impossible or that, alternatively, it is obvious that we can successfully pursue social goals. Here we shall see that Hayek was almost certainly correct about our inability to guide the system at the macro level, but there are other modes and levels of governance that need to be explored. On this matter, our inquiry leads to conclusions much more in tune with the work of the Ostroms than of Hayek. Part III concludes with an argument for the general superiority of democratic over expert-driven governance.

¹ Binmore, *Natural Justice*, p. 1.

² Kauffman, *A World beyond Physics*.

³ Rawls, *Justice as Fairness*, pp. 702. I have explored the idea of a neighborhood in some depth in the *Tyranny of the Ideal*, pp. 74ff.

⁴ Except, perhaps, through the use of great and oppressive coercion.

PART I

THE RISE OF A NORMATIVE
SPECIES

Are We Fit for the Open Society?

This Part takes up Hayek's first unsettling thesis: that our evolved, tribal, and egalitarian sentiments are in deep conflict with the impartiality and inclusiveness of the Open Society. Hayek, I shall argue, was correct that the core of human morality—which I shall call the “Modern Egalitarian Package”—arose during the long hunter-gatherer period in our history. However, Hayek largely overlooked the extent to which human cooperation arose on the basis of strong reciprocity, and the way in which our egalitarianism is manifested in impartial norms that protect against bullying and domination. These features of the Modern Egalitarian Package allow it to be scaled up to large impersonal moral networks.

§5 A Natural History of Moral Order

As we saw in the Prolegomenon, Hayek rejects the constructivist, abstract rationality-based, tradition of the social contract, and often looks to the Scottish Enlightenment and figures such as Adam Ferguson, who constructed natural histories of civil society.¹ The insight underlying the natural history approach is not simply the truism that institutions and societies have a history, but that the later can only be understood as an outgrowth of the earlier. When contemporary economists sing the praises of Adam Smith, they seldom focus on Book III of *The Wealth of Nations*—a natural history of opulence from the fall of Roman Empire. But Smith too holds that one cannot understand the later except in light of the developments and possibilities of the earlier. This renders the analysis of any period or institution more “evolutionary” and less strictly analytic.

In my view, Rousseau’s *Discourse on the Origins of Inequality* is one of the outstanding natural histories in political philosophy. This, alas, may seem outrageous to many of Hayek’s most loyal followers—he often takes Rousseau to be a villain in his narratives, an exemplar of so much of what was wrong with French thought.² Yet *The Discourse on Inequality* is a criticism of Hobbesian static analysis that an admirer of Hayek should appreciate. The competitive, nasty, creatures on which Hobbes focused, as well as the social contract they would accept, are, Rousseau argued, indicative of a period in humanity’s history, a result of a development that not only colors what they are, but also affects what they can be. Hayek, of course, tells a much different tale: a rise from a closed small group egalitarianism to a free open society—though again, a rise colored by its roots. Indeed, they agree that the right account begins with an egalitarianism that cannot be reclaimed, yet continues its allure. “Discontented with your present state . . . you will perhaps wish it were in your power to go backward.”³

My aim in this Part is to sketch a natural history of our moral order based on contemporary research. Strikingly, we shall find strong confirmation of Rousseau’s and Hayek’s insistence on the central role of our egalitarian past, and the rise of large-scale hierarchy that ended it.⁴ Like Hayek and Rousseau, our natural history is premised on the supposition that no time-slice picture of human social life does justice to its nature: we possess a matrix of social tendencies that can only be adequately understood by understanding its genesis and internal relations. Hayek was, we shall see, entirely correct that our moral nature was critically shaped in an extraordinarily egalitarian period. To understand this period and its implications for the Open Society, we need to understand both the materials on which it built and the moral architecture which it

¹ Ferguson, *An Essay on the History of Civil Society*. See also Book III of Smith, *Wealth of Nations*. For an overview, see Skinner, “Natural History in the Age of Adam Smith.”

² But see Hayek’s “The Political Ideal of the Rule of Law,” p. 143ff. for a more nuanced treatment of Rousseau.

³ Rousseau, *Discourse on Inequality*, p. 133. “Man has been civilized very much against his wishes.” Hayek, *The Political Order of a Free People*, p. 168.

⁴ For a contemporary support of Rousseau’s history, see Flannery and Marcus, *The Creation of Inequality*.

bequeathed to us. Only then can we adequately evaluate Hayek's unsettling thesis that we are in some ways unfit for the Open Society—discontented, and too often wishing it was within our power to go backward.

Hayek was fond of pointing out that the Scottish philosophers of the eighteenth century developed an account of social evolution long before Darwin's theory of natural selection: indeed, it was Darwin who borrowed from earlier social theory.⁵ Natural history is, however, an evolutionary explanation only in a broad sense: as I have said, it traces the development of, say, moral order on the assumption that the later can only be understood as outgrowth of features of the earlier.⁶ This, of course, is an essential commitment of all evolutionary explanation, but a full evolutionary explanation must provide the mechanisms by which more adaptive variants were selected over less adaptive variants and transmitted to the next generation. Without an analysis that explains diversity, selection, and transmission, there is not a fully adequate evolutionary explanation in the Darwinian sense. Rousseau's natural history did not endeavor to present such a complete account; he explicitly abstracted from specific mechanisms, allowing that the history he sketched "might have happened in various ways."⁷ For the most part (but see §11.4), I too shall prescind from debates about the various mechanisms, and, like Rousseau, focus on major evolutionary developments in which one type of social order gave way to another. I again take inspiration from Rousseau, charting a series of revolutions marking the ups and downs of egalitarian moral order.

§6 The "Starting Point"

6.1. Ancestral *Pan*

As we shall see, Hayek's claim that our moral nature critically evolved in small face-to-face groups has been confirmed by a large body of recent work. However, for the same reasons that we should dismiss the idea that contemporary humans are "blank slates" that can learn and conform to just about any morality,¹ we must remember that our hunting-gathering Pleistocene ancestors were not blank slates on which a thoroughly egalitarian or so-called tribal morality could be impressed. And we need to keep in mind Rousseau's criticism of Hobbes—that, in looking for "natural man," Hobbes did not go back far enough in history.² If we simply start with, say, the small scale

⁵ Hayek, *The Constitution of Liberty*, p. 59. At the heart of Darwinian natural selection is the Malthusian doctrine of a "struggle for existence" (the title of chap. 3 of *The Origin of Species*) produced by population growth outstripping the environment's carrying capacity. See Malthus, *An Essay on the Principle of Population*, p. 14.

⁶ On natural history as an "evolutionary" doctrine, see Hayek, "The Results of Human Action but Not of Design," p. 301n.

⁷ Rousseau, *The Discourse on Inequality*, p. 159.

¹ Pinker, *The Blank Slate*. Recall here the effort to create a "New Soviet Man," or B. F. Skinner's *Walden Two*.

² Rousseau, *Discourse on the Origin and the Foundation of Inequality among Men*, p. 216.

hunter-gatherer societies of the Late Pleistocene era, we run the risk of making precisely the error that Hayek warns against concerning the Great Society: just because it evolved does not mean that we are not ambivalent about it. Our moral nature need not be, as it were, optimized to its current social environment—and that was true in the Late Pleistocene era, too. If we wish to understand the evolution of human morality in small-scale bands, we must go back further and examine the ancestors of those who formed those bands. Of course, Rousseau was right on another point: no matter how far we go back, we will always find a creature with specific dispositions.³ There is no true “starting point” in evolutionary accounts. Rather, we commence with a moment in an already evolved history and see what evolved from that.

A good candidate for our starting point is “ancestral *Pan*,” the posited common ancestor of humans and our two closest primate relatives, chimps and bonobos. The human line diverged from chimpanzees and bonobos 5–7 million years ago; after that, a line evolved leading to the ultra-social *Homo sapiens*.⁴ Ancestral *Pan* was not a blank slate, a pure self-interested Machiavellian, or an impulsive wanton,⁵ but a primate that, like Rousseau’s “natural man,” was already endowed with capacities and tendencies that must have been the basis for much of what followed. Although our focus will be on ancestral *Pan*, it will also be helpful to sometimes make reference to an even more distant relative, the so-called Common Ancestor of gorillas, chimps, bonobos, and humans.⁶ The ancestral *Pan* line divided from the gorilla line 8–10 million years ago.

6.2. Behavioral Phylogenetics

To arrive at a more specific idea as to what *Pan* was like, I shall rely on “behavioral phylogenetics.” This method employs the principle of parsimony and assumes that the preferred explanation of evolutionary change is the most economical. If we find, say, that contemporary humans, bonobos, and chimpanzees share a genetically based behavioral trait such as hierarchical organization, we infer that the most recent ancestor shared by all three species possessed that trait and passed it to them.⁷ Of course, it *could* have been that (a) all the behaviors have genetic bases, but arose independently, or (b) they might not have any genetic basis at all, but are learned behaviors that are shared.⁸ But (a) would violate the assumption of parsimony in evolution, so it is not the preferred explanation (which is not to say that it could not be correct).

³ Ibid., p. 135.

⁴ The chimpanzee and bonobo lines diverged between 2 and 3 million years ago. In addition to humans, the other main ultra-social species are ants and bees. Humans, ants, and bees demonstrate an intensive form of cooperative social life marked by a division of labor in huge “colonies.” See Wilson, *The Social Conquest of the Earth*.

⁵ See de Waal, *Bonobo*, pp. 150–8.

⁶ See Boehm, *Moral Origins*, chap. 5; Boehm, *Hierarchy in the Forest*, pp. 152ff; Wrangham and Peterson, *Demonic Males*, chap. 2; de Waal, *Bonobo*, pp. 57–62.

⁷ Boehm, “Ancestral Hierarchy and Conflict,” p. 845. Citation omitted. See also Brosnan, “Nonhuman Species’ Reactions to Inequity,” p. 161; Wrangham and Peterson, *Demonic Males*, chap. 2.

⁸ Another possibility is that they have genetic bases but are simply the result of similar random accumulated changes. We focus here on central features to their way of life that are most plausibly seen as functional adaptations.

To minimize the chance of (*b*), we focus on behaviors basic to the social life of species that live in very different environments (as chimps and humans do); in this case, we may have some confidence that these are not merely independently learned behaviors.

On Christopher Boehm's view, the

[t]wo main rules for reconstruction help to keep the analysis conservative. First, if a major behavior pattern is found unanimously and unambiguously in bonobos, chimpanzees, and human foragers and if no evolutionary evidence exists to the contrary, the probability of its having been present in their (quite recent) shared ancestor will be very high. Second, I suggest that, if among the three extant species a trait is expressed strongly by two but weakly by the other, the ancestral assessment should be based just on the least common denominator.⁹

This, then, is a conservative method, biased toward false negatives rather than false positives. Only if all three existing descendants share a behavior do we attribute it to *Pan*, but of course it could be that *Pan* had the trait and one of the existing species lost it (or by the second rule, it significantly weakened in one). Thus, we would have a false negative whenever a descendant species lost a feature of the common ancestor, a reasonably likely event. False positives would occur when in fact the same trait in the descendant species had arisen independently or was simply the result of convergent independent social learning. In evolution the same trait (say, flying) can arise by an independent route in different species (insects, birds);¹⁰ this possibility, however, is minimized when the species are closely phylogenetically related, as are chimps, bonobos, and humans. And again, the more the species live in very different environments, the more we can be confident that the behavior is not the result of convergence in learning.

6.3. Ancestral *Pan*: A Plastic Hierarchy

Alpha Despotism and Coalitions

Rousseau's *Discourse* is the itinerary from natural equality to modern inequality, of egalitarian paradise lost. Was *Pan* egalitarian or despotic?¹¹ Three factors seem relevant in determining the despotic or egalitarian nature of a social group: (1) the extent of competition in resource allocation (e.g., food, mating); (2) the existence of social

⁹ Boehm, "Ancestral Hierarchy and Conflict," p. 845. Citations omitted.

¹⁰ Complications are hidden in this seemingly clear statement. The distinction between "homologies" and "analogies" describes relational properties among a certain phylogenetic group, and not invariant descriptions. For example, looking at vertebrates, the wings of bats and birds are both developments of the forelimb, and can be considered as homologous (so they did not arise independently), while in relation to the mammal/avian distinction, their wings are considered analogous, having arisen independently within those classifications. See Preuschoft and van Hooff, "Homologizing Primate Facial Displays," p. 122. These technicalities can be set aside here: we are only concerned with traits and their development within this group of three primates.

¹¹ The categorizations have been formally modeled in evolutionary theory. See Vehrencamp, "A Model for the Evolution of Despotic versus Egalitarian Societies."

hierarchies that allow alphas (male or female) to dominate others; and (3) the presence of political coalitions to secure alpha control and social order.¹² On all these traits, chimpanzee societies score high on despotism. Alpha males dominate the entire group, and males generally dominate females.¹³ Physical displays of superiority and enforcement of superiority by blows to lower ranked members, from the mild to severe, are common.¹⁴ Dominance is not based only on strength, but depends as well on political alliances with other males and females. De Waal recounts the extended political struggle between two chimps, Yeroen and Luit, which eventually led to Yeroen's downfall. In the face of Luit's challenge, Yeroen sought to rally the females to his cause, while Kikkie, a young male, joined on Luit's side, seeking to neutralize the females' support of Yeroen.¹⁵ Luit eventually emerged as alpha, only to be subsequently dethroned by a Kikkie-Yeroen coalition, with Kikkie as the new alpha. "If we follow Harold Laswell's [*sic*] famous definition of politics as a social process determining 'who gets what, when, and how,'" de Waal points out, "there can be little doubt that chimpanzees engage in it"¹⁶—and a despotic form of it.

Bonobos also tend to despotism, having a clear social hierarchy.¹⁷ The critical difference is that males are not dominant over females. Some hold that the sexes are "codominant,"¹⁸ while others observe a female dominance.¹⁹ A critical difference between chimps and bonobos is the nature of their political life. While bonobo males are, like chimpanzees, larger than females, females join together in coalitions with other females, but (unlike chimps) male-male coalitions do not occur, leading to a female-dominated politics.²⁰ A male's participation in politics takes the form of coalitions with his mother and her allies, and his alpha male status, which seems associated with higher food access, requires the support of his mother.²¹ However, bonobo males are not nearly so obsessed with the grand game of primate politics, though neither are they the sort of New Age primate some have seen them as—males show definite indications of bite marks, perhaps inflicted by females in the face of challenges to their authority.²²

The real question concerns humans, who of course display the entire gamut of social relations, from resolutely egalitarian to absolutely despotic. Rousseau would have it that we started out as natural egalitarians and became corrupted by power and oppression. And it is true that the dominant view is that humans generally lived a highly egalitarian lifestyle with an egalitarian politics during the Late Pleistocene era (§7).²³

¹² Boehm, *Hierarchy in the Forest*, p. 2.

¹³ For both sexes, those who are dominant have better access to food, while dominant males have more opportunities for mating. *Ibid.*, p. 26.

¹⁴ De Waal, *Chimpanzee Politics*. See also Boehm, *Hierarchy in the Forest*, pp. 22ff.

¹⁵ De Waal, *Chimpanzee Politics*, chap. 2. Boehm describes chimpanzee alphas as more akin to bullies than human despotic rulers; *Hierarchy in the Forest*, p. 26.

¹⁶ Preface to the 1998 edition of *Chimpanzee Politics*. See also Boehm, *Hierarchy in the Forest*, pp. 22ff.

¹⁷ Boehm, *Hierarchy in the Forest*, pp. 30, 133–7.

¹⁸ Wrangham and Peterson, *Demonic Males*, p. 205; Boehm, "Ancestral Hierarchy and Conflict," p. 844.

¹⁹ See de Wall, *Bonobo*, pp. 75ff.

²⁰ See Boehm, "Ancestral Hierarchy and Conflict," p. 844.

²¹ *Ibid.*; Boehm, *Hierarchy in the Forest*, pp. 133–7, 155–6.

²² Boehm, *Moral Origins*, p. 107.

²³ Gintis, van Schaik and Boehm, "Zoon Politikon."

One of Boehm's significant contributions has been to show that Rousseau's egalitarianism was something of an intermediate stage, between the despotism of ancestral *Pan* and the Holocene era's more hierarchically organized humans. On his reading of the evidence, our Common Ancestor (which, recall, includes gorillas),

lived a social life that was heavily determined by individual tendencies to dominate and, ambivalently, submit. . . . if we look closely at gorillas, bonobos, chimpanzees, and humans, we'll see a noteworthy shared tendency for alpha males to appear at the tops of pecking orders, and, linked to the predictable and strong competition for high rank that goes with this. . . . We only have to think of a modern Hitler or Mao or Stalin as an alpha male at the top of a hierarchy—or of an American president as a less all-powerful alpha who nonetheless gives orders to a huge military—and it's obvious that we share the apes' hierarchical tendencies and have a strong potential to develop alpha males.²⁴

Thus, as we shall see (§7.3), there is strong evidence that Late Pleistocene egalitarianism was a systematic attempt to control alpha male behavior, not the "natural staring point" of primal egalitarianism.

Resistance to Alpha Domination

Boehm insists that submission to dominants was likely "reluctant" in the Common Ancestor: gorillas, chimps, and bonobos all display significant efforts of subordinates to neutralize alpha despotism.²⁵ At least in contexts where females chimps have close contact (such as zoos), they can act in support of, or opposition to, a high-ranking male, sometimes helping to depose a disliked one.²⁶ Female coalitions, of course, are most characteristic of bonobos: in the wild as well as in zoos, coalitions of females control alpha males' attempts to dominate. "All available observations tell the same story. Female bonobos turn the tables on males. And if bonobo males throw their weight around and become overly aggressive, they are liable to be suppressed by females."²⁷ De Waal reports a striking instance where a high-ranking male sought to jump on a branch with a low-ranking female and her infant, threatening the infant. When they fell to the ground, fifteen or more adult bonobos attacked the high-ranking male and dragged him about.²⁸

Boehm goes so far as to argue that Ancestral *Pan* displayed the roots of a concern for personal autonomy—"a strong distaste for being intimidated and bossed around"²⁹—combined, we might add, with a distinct taste (at least on the part of some) for bossing

²⁴ Boehm, *Moral Origins*, pp. 95–6.

²⁵ *Ibid.*, p. 96. See also de Waal, *Chimpanzee Politics*, p. 122, who points out the great reluctance of deposed alphas to submit to the new ruler.

²⁶ Boehm, *Moral Origins*, p. 111.

²⁷ Wrangham and Peterson, *Demonic Males*, p. 209.

²⁸ De Waal, *Bonobo*, p. 22.

²⁹ Boehm, *Moral Origins*, p. 151.

others around. It should be stressed that this ambivalence about hierarchy is by no means inevitable in a species whose social life is organized hierarchically. The other great ultra-social species, ants and bees, are organized by rigid classes and a social hierarchy that determine mating opportunities but are entirely without subordinate rebellions.³⁰

Strategic Reasoning

Chimpanzees and bonobos are typically described as “Machiavellian” cooperators.³¹ As Michael Tomasello concludes, “chimpanzees and bonobos are built for competition. Thus, not only are they intentional, decision-making agents, who make instrumentally rational decisions themselves, but they also perceive others as intentional, decision-making agents with whom they must compete.”³² Thus “chimpanzees and bonobos live their lives embedded in constant competition for resources, so they are constantly attempting to outcompete others by outfighting them, outsmarting them, or outfriending them.”³³ Success in hierarchical social groups requires important cognitive abilities, including mind-reading and most probably an ability to deceive others.³⁴ Strikingly, experiments indicate that chimps perform better than adult human subjects in some strategic contexts.³⁵ Bonobos have not been as well studied as chimps, but male bonobos seem to share many of the traits of male chimps, competing and scheming their way to alpha status.³⁶ On top of this, though, bonobo social life is built on mother-son alliances and alliances among females to control males.

Proto Rule-Governance

There is, then, good reason to suppose that Ancestral *Pan* led a complex social life in which competition for alpha status was sometimes matched by the rebellion of the lower ranked. Survival in this politically fluid social hierarchy required significant strategic reasoning. After all, without strategic reasoning, “politics” is simply brute conflict. However, while the hierarchical structure possessed a political fluidity, it also

³⁰ To paraphrase Darwin (*The Descent of Man*, p. 473), if humans had evolved as bees did, subordinates would think it their highest moral duty never to rebel.

³¹ See Byrne and Whiten, “Machiavellian Intelligence.”

³² Tomasello, *A Natural History of Human Morality*, p. 21.

³³ *Ibid.*, p. 26.

³⁴ Byrne and Whiten, “Machiavellian Intelligence,” pp. 4–11; Laland, *Darwin’s Unfinished Symphony*, p. 135. See also de Waal, *Chimpanzee Politics*, p. 128; Tomasello, *A Natural History of Human Morality*, p. 26.

³⁵ In a fascinating recent series of experiments, Martin et al. had chimpanzee mother-child pairs playing competitive matching pennies games. Each player puts out a “penny” (decides to play either X or Y). Player A wins if the pennies are both heads or both tails (the matcher), the other player wins if they are different (the mismatcher). The best response of players in these games is to play a Nash mixed strategy, attempting to counter the other player’s move. Martin et al. found that chimps are much better at finding the correct mixed response. “These results suggest that the reason the chimpanzees converge more sharply to mutually best-responding (i.e., are closer to Nash equilibrium) is because they adjust to opponent behavior and to changes in incentives more strongly.” Martin et al., “Chimpanzee Choice Rates in Competitive Games,” pp. 2–3.

³⁶ Wrangham and Peterson, *Demonic Males*, pp. 210ff.

provided a relatively stable social structure based on what we might call the “proto-rules” of rank: “[e]veryone knows these social rules” based on rank—even when the aim is to circumvent them.³⁷

Denise Cummins has argued that the origins of deontic reasoning lie in hierarchical social living: “remaining and surviving within a dominance hierarchy depends crucially on the capacity to detect and respond appropriately to permissions, obligations, prohibitions, promises, threats and warnings. Failure to do so carries a high risk of inciting agonistic encounters or ostracism.”³⁸ Hierarchies *are* a set of enforced social rules. The typical philosophical move of deriving rule-following behavior from strategic behavior is not only ill-conceived concerning humans, but appears to misunderstand *Pan*, who already had a complex social life composed of at least proto-rule following *and* strategic reasoning, seeking both to exploit and to circumvent the rules of hierarchical social life. In chimpanzee society, rule violators—for example, lower status males who seek to mate with higher status females—are apt to be punished by the alpha male, yet lower status males often seek to violate the rules when they are unobserved. To know what to do requires that one knows one’s place, and the obligations and permissions that attach to it—even if one sometimes seeks to avoid them. “At a bare minimum, humans and the other ape species that originated in Africa . . . know intuitively what a ‘rule’ is. This usually involves a stronger individual’s demanding a certain type of behavior, and the expectation is backed up by a potentially hurtful authority.”³⁹

6.4. Ancestral *Pan*: Other Traits

These four traits—despotism, resistance to domination, strategic reasoning and proto-rule-governance—are plausibly seen as core features of *Pan*, intimately related to its distinctive form of social organization. “Natural man” was rather more complicated than Rousseau’s solitary primate. These traits are significant in all three of the existing species that descended from *Pan*. Chimps seem strongest on despotism, followed by bonobos and humans; humans seem strongest on resistance to domination (more on that anon), with chimps likely the weakest, while chimps and humans score high on strategic reasoning. And, as we shall see, humans score highest on rule-following. Consistent with the second tenet of behavioral phylogenetics, the strength of the traits we attribute to *Pan* is its weakest manifestation in any of the three descendent species. All four of these traits would appear to have been critical features of *Pan* and were part of its distinctive form of social life.

³⁷ Ibid., p. 169. I say “proto-rules” as there is no need to assume the full set of rule-based competencies characteristic of humans. Remember, our aim is to see the competencies on which human evolution built, not to claim that human competencies were fully present in our ancestors.

³⁸ Cummins, “Evidence for the Innateness of Deontic Reasoning,” p. 166.

³⁹ Boehm, *Moral Origins*, p. 106.

Territoriality and Xenophobia

To some extent, ancestral *Pan* was also almost certainly territorial and xenophobic.⁴⁰ Here, however, we observe great variation among our three species, and so we should be especially alert to assuming only the lowest common denominator. Chimps are highly territorial and xenophobic. They have been known to patrol the borders of their territory, and engage in deadly raids against neighboring groups when they have the numerical advantage.⁴¹ Like humans, chimpanzees sometimes expand their territory by violently displacing neighboring groups through systematic raiding. Contact between different chimpanzee groups almost always leads to conflict, sometimes deadly. In contrast, while bonobos (especially males) are apprehensive when first meeting other groups, violence is not a typical outcome.⁴²

Hunting and Sharing

One way in which *Pan* appears to have clearly departed from the Common Ancestor (from which our three species, as well as gorillas, evolved) is meat eating and hunting. While gorillas are vegetarians, chimpanzees hunt smaller game, such as monkeys. Because this often involves catching them in the three-dimensional space of the forest canopy, small groups often coordinate.⁴³ Compared to chimps, bonobos hunt less, but certainly are predators.⁴⁴ Like chimps, they appear to sometimes hunt in groups, but overall seem to hunt individualistically rather than collectively.⁴⁵ Anyone who has lived in northern Minnesota can testify that humans too have been observed to hunt.

Chimps and bonobos appear to have similar though somewhat complex practices of meat possession and sharing among adults.⁴⁶ Sometimes a lone chimp catches a small animal and, regardless of status, may be allowed to keep it for himself. However, chimps do have a distinctive scream that seems to indicate their willingness to share meat.⁴⁷ In most cases, alpha and higher status males and allies get first dibs and often expropriate the carcasses of other hunters;⁴⁸ pregnant females get more than others. There is debate whether this process is mostly one of tolerated theft by lower status members, begging by lower status chimps, or reinforcing alliances.⁴⁹ A recent

⁴⁰ Ibid., p. 135.

⁴¹ Wrangham and Peterson, *Demonic Males*, chap. 4; McGrew and Feistner, "Two Nonhuman Primate Models for the Evolution of Human Food Sharing," pp. 231–2.

⁴² De Waal, *Bonobo*, p. 65. See also Wrangham and Peterson, *Demonic Males*, p. 214.

⁴³ De Waal, *Bonobo*, p. 124; Stanford, *The Hunting Ape*, pp. 3ff.

⁴⁴ Ibid., p. 65. Wrangham and Peterson hypothesize that because chimpanzees, but not bonobos, competed with gorillas for non-meat food sources, they evolved to obtain a larger share of their calories from meat. *Demonic Males*, chap. 11.

⁴⁵ Boehm, *Moral Origins*, p. 136. Cf. de Waal, *Bonobo*, p. 65.

⁴⁶ As we might expect, sex differences have less impact on food sharing among bonobos. McGrew and Feistner, "Two Nonhuman Primate Models for the Evolution of Human Food Sharing," p. 232.

⁴⁷ De Waal, *Bonobo*, p. 193; McGrew and Feistner, "Two Nonhuman Primate Models for the Evolution of Human Food Sharing," p. 232.

⁴⁸ Wrangham and Peterson, *Demonic Males*, p. 216; Stanford, *The Hunting Ape*, p. 194.

⁴⁹ McGrew and Feistner, "Two Nonhuman Primate Models for the Evolution of Human Food Sharing"; Boehm, *Moral Origins*, pp. 138–41.

experiment by Melis, Hare, and Tomasello found only a very weak tendency of chimps to preferentially help those who had just helped them.⁵⁰ And although there has been dispute on this matter, it now seems doubtful that chimps or bonobos have any sense of fair division.⁵¹ Meat sharing seems ultimately motivated by a mix of power and sympathy.⁵² As Boehm concludes, “[i]n both bonobos and chimpanzees, the impression is one of gluttonously selfish possession, combined with rampant cronyism when it comes to limited sharing of this most precious of all foods. The overall sharing process seems to be shaped both by individual political power and by personal alliances.”⁵³

Reasoning and Social Learning

We have seen (§6.3) that chimps reason strategically, and we inferred that *Pan* probably had such abilities. This raises the question of the general intelligence of *Pan*. It has been estimated that the human brain has approximately quadrupled in size over the last three million years,⁵⁴ and it might reasonably be supposed that this indicates a tremendous increase in cognitive abilities across the board. On this view, we would expect that compared to contemporary humans, *Pan* would score much lower across a wide range of cognitive abilities. Yet it does not seem that the data really fit this natural assumption.⁵⁵ In an important experiment, Herrmann and her colleagues compared three groups of subjects on a range of cognitive tasks: 106 adult chimpanzees (3 to 21 years of age, with a median of 10 years), 32 orangutans (3 to 10 years of age, with a median of 6), and 105 humans (aged 2.5).⁵⁶ The rationale behind comparing adult primates with human toddlers was to isolate humans’ natural cognitive abilities from the influences of cultural knowledge: after roughly three years, children begin to quickly amass cultural knowledge. Hermann et al. tested all three groups on tasks in what they described as the “physical” domain (spatial, quantitative, and causal skills), and the “social” domain (social learning, communication, and theory of the mind). They found that in both domains, chimps and humans outperformed orangutans.⁵⁷ However, within the physical domain, human and chimp subjects did, overall, about equally well. The great difference was in the social domain: human toddlers were manifestly better at social skills.⁵⁸

⁵⁰ Melis, Hare, and Tomasello, “Do Chimpanzees Reciprocate Received Favours?”

⁵¹ See Tomasello, *A Natural History of Morality*, pp. 32–4. For evidence of “inequity aversion” in chimps, see Brosnan, “Nonhuman Species’ Reactions to Inequity and their Implications for Fairness”; Bierhoff and Rohmann, “Conditions for Establishing a System of Fairness.”

⁵² On sympathy, see Tomasello, *A Natural History of Morality*, pp. 28ff.

⁵³ Boehm, *Moral Origins*, p. 138.

⁵⁴ Laland, *Darwin’s Unfinished Symphony*, p. 26.

⁵⁵ There is, of course, dispute about this matter. For a view more sympathetic to the general intelligence hypothesis, see *ibid.*, chap. 6.

⁵⁶ Herman et al., “Humans Have Evolved Specialized Skills of Social Cognition.”

⁵⁷ Orangutans are human’s most distant relatives in the great ape family, splitting from the Common Ancestor approximately 11–15 million years ago. In many ways these solitary creatures are most like Rousseau’s “natural” humans.

⁵⁸ In a subsequent study, Herman and her coworkers found that chimp cognitive abilities factored into the two domains in this study, physical and social, but human abilities best fit a three-factor model, of spatial, physical, and social domains. “The Structure of Individual Differences in the Cognitive Abilities of Children and Chimpanzees.”

These results are consistent with important problem-solving experiments with chimps and human children.⁵⁹ In these experiments, human and chimp subjects observed a human demonstrator use a tool to retrieve a reward from a puzzle-box. In one treatment, the human demonstrator engaged in some instrumentally effective actions in solving the problem, as well as ones that the learner could see were ineffective. While chimpanzees tended to copy only the instrumentally effective demonstrations, human children were more likely to emulate the entire demonstration, ineffective as well as effective. A strong tendency for social learning—doing precisely what successful others do—appears to be a distinctively human trait (§11). This strong tendency for humans to emulate each other may well be important in explaining why, in games like matching pennies, human mismatchers underperform if they can pick up cues that indicate what the other player is about to do. Rather than playing the opposite, at least initially there is a tendency to mimic the other player.⁶⁰

None of this is to deny that chimps engage in social learning. Evidence from chimps in the wild indicate that they do indeed possess cultural knowledge. For example, some groups of chimps use tools to crack open nuts, and among those who do there is group variation in the type of tools used: at one location the chimps employ stone hammers and anvils, while in another wooden hammers are used as well. Recent analysis indicates that neither genetic variation nor different environmental conditions can explain these differences: there seems to be genuine social learning within different groups.⁶¹ Overall, it seems safe to conclude that *Pan* had considerable cognitive skills and engaged in rudimentary social learning.⁶²

§7 The Egalitarian Revolution

7.1. The Remarkable Transformation

As I have said, Rousseau's *Discourse* traces the rise of inequality through a series of revolutions—the development of property, political authority, and, finally, despotism. For Rousseau, the rise of inequality was a steady climb up (or, for him, steady descent). I too shall trace a series of revolutions, but we have seen that our “natural starting point” is not an idyllic state of equality, but the complicated dynamic

⁵⁹ Horner and Whiten, “Causal Knowledge and Imitation/Emulation in Chimpanzees (*Pan troglodytes*) and Children (*Homo sapiens*).” See also Call, Carpenter, and Tomasello, “Copying Results and Copying Actions in the Process of Social Learning.”

⁶⁰ Belot, Crawford, and Heyes, “Players of Matching Pennies Automatically Imitate Opponents' Gestures against Strong Incentives.” See also Cook et al., “Automatic Imitation in a Strategic Context: Players of Rock—Paper—Scissors Imitate Opponents' Gestures.”

⁶¹ Lycett, Collard, and McGrew, “Phylogenetic Analyses of Behavior Support Existence of Culture among Wild Chimpanzees.” See also Stanford, *The Hunting Apes*, chap. 3.

⁶² Herman et al., “Humans Have Evolved Specialized Skills of Social Cognition,” p. 1536.

hierarchical life of *Pan*. For us, the first human revolution was the *rise* of egalitarianism that characterized the Late Pleistocene period. We can now appreciate just how remarkable a rise that was. From a despotic, rather Machiavellian, species, Late Pleistocene humans evolved into highly egalitarian ultra-social creatures. *Homo sapiens* are indeed *ultra-social*, and in that sense we are very much like ants and bees. This is no exaggeration: ants, bees, and humans are the main “ultra-social species.” All three species are characterized by an intense division of labor, in which the goods needed for everyone’s existence are primarily socially produced.¹ In many ways, it is not remarkable that insects such as bees and ants achieve such intense social cooperation—the members of a hive are closely genetically related. Thus, it is typically thought that kin altruism explains the ultra-sociality of ants and bees.² Genetically, ant colonies and beehives are often considered a single organism. But humans evolved in groups with low genetic relatedness among competitive agents. Understandably, recent work on the evolution of human egalitarian sociability has focused on explaining how such a highly cooperative species evolved from hierarchical, competitive, *Pan* through natural selection. A variety of mechanisms and models have been put forward—among them mutualism,³ direct reciprocity,⁴ social selection,⁵ group (multilevel), selection,⁶ and gene-culture coevolution.⁷ No doubt more than one mechanism—perhaps all of them—were involved at various stages. And, of course, critical to any full evolutionary explanation is not only the mechanism, but an account of the adaptive challenges and opportunities that humans faced at various stages of our evolution. Adaptive pressures for extensive cooperative

¹ Hobbes recognized both the similarity and the difference: “It is true, that certain living creatures (as bees and ants), live sociably one with another, (which are therefore by Aristotle numbered amongst political creatures), and yet have no other direction, than their particular judgments and appetites, nor speech, whereby one of them can signify to another, what he thinks expedient for the common benefit; and therefore some man may perhaps desire to know, why mankind cannot do the same.” *Leviathan*, p. 108.

² If *X* and *Y* are different phenotypes that share most of their genes, then, from the “gene’s eye view,” it can be adaptive for *X* to act in a fitness-decreasing way to benefit *Y*. Famously, W. D. Hamilton generalized the conditions for selection by inclusive fitness. Supposing that organism *X* can provide adaptive benefit *b* to *Y* at an adaptive cost *c* to itself, then, if *r* is the degree of genotype relatedness between *X* and *Y*, this act of inclusive fitness or kin altruism promotes the interests of the genotype when: $rb > c$. In the absence of inbreeding, in humans the “coefficient of relatedness” (*r*) is .5 between parents and their children, .25 between grandparents and grandchildren, .5 between siblings, and .125 between first cousins. The *r* values fall quickly away after first cousins. See Hamilton, “The Genetical Evolution of Social Behaviour: I.” For a dissenting analysis, see Nowak, Tarnita, and Wilson, “The Evolution of Eusociality.”

³ See, e.g., Tomasello, *A Natural History of Human Morality*, pp. 13ff; Sterelny, “Cooperation, Culture, and Conflict,” pp. 43–6.

⁴ See, e.g., Binmore *Natural Justice*, pp. 79–82. Cosmides and Tooby, “Cognitive Adaptations for Social Exchange,” pp. 170–9.

⁵ Social selection is based on reputations for cooperation and reciprocity, and is thus a version of “indirect reciprocity.” See Boehm, *Moral Origins*, chap. 7; Alexander, *Human Social Evolution*, chap. 9; Nowak and Highfield, *Super Cooperators*, chap. 2. Social selection employs, among other mechanisms, sexual selection—the preference of females for more cooperative males. On the overlooked importance of sexual selection, see Miller, *The Mating Mind*.

⁶ See, e.g., D. S. Wilson, *Does Altruism Exist?*; Haidt, *The Righteous Mind*, chap. 9; Bowles and Gintis, *A Cooperative Species*; E. O. Wilson, *The Social Conquest of the Earth*; Boyd and Richerson, *The Origin and Evolution of Cultures*, chap. 11.

⁷ See, e.g., Richerson and Boyd, *Not by Genes Alone*; Henrich, *The Secret of Our Success*; Henrich and Henrich, *Why Humans Cooperate*, chap. 2.

hunting, social “nesting,” and success in inter-group conflict are among the many factors that have been proffered.⁸

As fascinating and important as these analyses are, as I stressed in section 5, our concern is more with the development from *Pan* to us—a natural history of human egalitarian morality. We thus seek to understand Late Pleistocene egalitarian social life in light of the traits of *Pan* from which it evolved, allowing that this evolution could have been produced by different mechanisms and probably by a complex mix. The assumption underlying our natural history is that we entirely misunderstand our egalitarian past if we do not see it as both built on—and seeking to cope with—the already evolved traits of *Pan*.

7.2. LPA Egalitarianism

We have good reason to conclude that modern, Late Pleistocene, humans lived in groups of between twenty and thirty,⁹ obtained a high percentage of their calories from hunting or fishing, and engaged in egalitarian meat sharing. If, however, we wish to make richer inferences about their social organization, we must make an additional assumption: that some contemporary hunter-gatherer societies approximate the social orders characteristic of the Late Pleistocene era. In his important study of contemporary Late Pleistocene appropriate (LPA) foraging societies, Boehm eliminated from consideration societies that have been heavily influenced by Western and market societies, those with some agriculture, those that trade with agricultural groups, those that rely on domesticated horses, and so on, ultimately identifying 150 contemporary forager societies whose way of life corresponds to what we know of Late Pleistocene hunter-gatherer bands. The critical assumption is that detailed analysis of these LPA societies allows us to make inferences about the social norms and core social concerns of our Late Pleistocene ancestors.¹⁰

This assumption is certainly not uncontroversial. It is generally thought that by 40,000 years ago humans were culturally and genetically modern, though some recent analyses suggest greater genetic change in the last 10,000 years.¹¹ Contemporary LPA foraging societies exist in the Holocene era of milder climates and arguably greater ease, or at least less uncertainty, in obtaining food. In the harsh and wildly variable Late Pleistocene climate,¹² it could well have been far less rare for groups to have faced

⁸ On hunting, see Stanford, *The Hunting Ape*; Boehm, *Moral Origins*, pp. 150ff. On nesting, see E. O. Wilson, *The Social Conquest of the Earth*; Henrich, *The Secret of Our Success*, pp. 298ff. On conflict, see Bowles and Gintis, *A Cooperative Species*, chap. 8; Henrich, *The Secret of Our Success*, chap. 10. Of course, there were many adaptive challenges and responses in the long evolution from *Pan* to the Egalitarian Revolution. For a helpful discussion, see Haidt, *The Righteous Mind*, pp. 146ff.

⁹ Friedman suggests a much larger figure of 150, with much larger numbers when groups fused. *Morals and Markets*, p. 16. See also Rose, who mentions 200 as the typical size of the groups in which humans evolved; *The Moral Foundations of Economic Behavior*, chap. 3. Closer examination shows that group size may be understood differently: average band size may differ from typical group size. See Bowles and Gintis, *The Cooperative Species*, p. 95.

¹⁰ Boehm, *Moral Origins*, pp. 78–82.

¹¹ Cochran and Harpending, *The 10,000 Year Explosion*; Turchin, *Ultrasociety*.

¹² See Richerson and Boyd, “Rethinking Paleoanthropology.”

such dire circumstances that sharing broke down, leading to the group splintering into family-sized, rather than band-sized units, with very different evolutionary dynamics.¹³ Nevertheless, the social organization of these societies corresponds to much of what we know about Late Pleistocene bands—they are mobile, stress sharing rather than storing meat, combine hunting with foraging, and live in core bands of twenty to thirty persons.¹⁴ And they often live in the harshest of contemporary environments. With care, we can draw useful inferences from the organization of LPA societies to form a richer idea of life in the sorts of bands in which humans evolved.

7.3. The Egalitarian Ethos of LPA Societies

A central feature of LPA societies is certainly equalized meat sharing. In these societies, meat is typically a highly prized and precious good, the distribution of which has great impact on the well-being of members (recall their *Pan* heritage, §6.4). In some groups and in some cooperative activities, something like strict equality holds; however, departures from equality are also observed—for example, kin-bias, biased shares on the basis of past behavior, as well as work effort.¹⁵ It is plausible to understand egalitarian sharing norms as having two core social functions.

First, and most obviously, they serve a means of variance reduction in food intake. LPA societies devote great energy to hunting, and highly prize meat. Hunting—as opposed to foraging—is a rather hit-and-miss affair; sometimes hunters come home with more than enough, other times not quite enough, and other times nothing at all.¹⁶ Managing this variance is a general problem for all social carnivores. Like *Pan*, most social carnivores handled it through dominance hierarchies; those at the top leave meat for others after taking their share. Only in humans does variation reduction take place via equalization.¹⁷

The intimate connection between human egalitarianism and meat eating and sharing is quite clear, though we cannot be certain that it was hunting that drove the egalitarianism. Mary Stiner and her colleagues discovered differences in the bones of the carcasses of human kills between 400,000 and 200,000 years ago at Qesem Cave in Israel.¹⁸ Bones from carcasses from 400,000 years ago demonstrate that the human hunters employed tools to cut the meat, but the cut marks indicate the presence of a number of different cutting implements employed at different angles. Evidence from this earlier period suggests that meat distribution systems were less staged or canalized than those typical of Middle Paleolithic, Upper Paleolithic, and later humans. The evidence for

¹³ Boehm, *Moral Origins*, pp. 274ff. On the other hand, it could well have been such instability that increased the benefits of cooperation. See Bowles and Gintis, *The Cooperative Species*, pp. 93ff.

¹⁴ This is on the low end of many estimates (see note 9 of this section), but band size of thirty is compatible with larger groups who, for example, share bride networks and trade.

¹⁵ See Kaplan and Gurvan, “The Natural History of Food Sharing,” pp. 102–3. See also Stiner, “Carnivory, Coevolution, and the Geographic Spread of the Genus *Homo*.”

¹⁶ Flannery and Marcus, *The Creation of Inequality*, p. 32.

¹⁷ Boehm, *Moral Origins*, pp. 142–3.

¹⁸ Stiner, Barkai, Gopher, and O’Connell, “Cooperative Hunting and Meat Sharing 400–200 KYA at Qesem Cave, Israel.”

procedural interruptions and diverse positions while cutting flesh at Qesem Cave may reflect, for example, more hands (including less experienced hands) removing meat from any given limb bone, rather than receiving shares through the butchering work of one skilled person. Several individuals may have cut pieces of meat from a bone for themselves, or the same individual may have returned to the food item many times. Either way, the feeding pattern from shared resources may have been highly individualized, with little or no formal apportioning of meat, something we would expect of *Pan*. Kills from 200,000 years ago display much more uniform cut marks, indicating a single cutter, who cut and distributed the kill. A very plausible hypothesis is that by this time humans were, or were well on their way to becoming, distinctly egalitarian hunters. In contrast to what we can infer from the cuts of 400,000 years ago, distribution of the kill does not seem determined by competition among the hunters (where we can suppose the more dominant took the best, first), but by a designated cutter allocating shares of the kill. To be a bit more speculative, it looks as if the socialized primate carnivores of 400,000 years ago were becoming egalitarian hunters by 200,000 years ago. It is very difficult not to conclude that egalitarian sentiments had already taken root by this period. Assuming modern humans had appeared by 45,000–40,000 years ago, there were 6,000–8,000 generations for egalitarian sentiments to evolve from what we can infer was their first appearance, somewhere between 250,000 and 200,000 years ago. It seems generally accepted that, in marked distinction to our primate cousins, humans are deeply attuned to considerations of fairness.¹⁹

It may well be that the *second* function of egalitarianism is critical here: suppression of assertions of dominance. We have seen that two of the core features of *Pan* were alpha dominance and subordinate resistance (§6.3). A fundamental feature of LPA egalitarianism seems to be highly successful subordinate control of alpha dominance. Indeed, Boehm argues that the LPA egalitarian revolution *was* essentially a successful effort by subordinates to control would-be bullies and upstarts.²⁰ As Boehm describes them, the truly fundamental feature of LPA societies is resolute and sustained suppression of would-be dominant members, and this most definitely includes would-be dominant hunters. Nomadic hunter-gatherers, Boehm concludes, are universally “and all but obsessively” concerned with resisting would-be dominators and bullies. Thus, he concludes, forgers are not concerned with absolute equality of outcomes, but equalization of shares as a way of resisting attempts by would-be dominant members to push them into a subordinate role.²¹ “Minimally, this means that all the active hunters (generally the adult males) insist on being seen as equal and that among themselves they tolerate no serious domination—be this in hogging vital food resources or in bossing others around.”²² Boehm’s claim, then, is that this is the type of social structure under which our egalitarian sentiments evolved—one based on a constant monitoring of would-be dominators and bullies.

¹⁹ Laland, *Darwin’s Unfinished Symphony*, p. 20. But cf. Henrich, Heine, and Norenzayan, “The Weirdest People in the World?”

²⁰ Boehm, *Hierarchy in the Forest*, pp. 172ff. Boehm’s analysis of hunter-gatherer societies is widely accepted. See, for example, Tomasello, *The Nature of Morality*, p. 43; Haidt, *The Righteous Mind*, pp. 198ff, 276; Flannery and Marcus, *The Creation of Inequality*, p. 59; Turchin, *Ultrasociety*, pp. 103ff, 163ff.

²¹ Boehm, *Hierarchy in the Forest*, p. 68.

²² Boehm, *Moral Origins*, p. 109.

In this environment, social sanctions are applied to those who cannot resist attempting to bully or subordinate their fellows, or who even go too far in self-praise. Norms against display of dominance through hunting are typically deeply engrained. Consider the report of Richard Borshay Lee's !Kung informant:

Say that a man has been hunting. He must not come home and announce like a braggart, "I have killed one in the bush." He must first sit down in silence until someone comes up to his fire and asks, "What did you see today?" He replies quietly, "Ah, I'm not good for hunting. I saw nothing at all ... maybe just a tiny one." Then I smile to myself because I now know that he has caught something big.

As another member of the group says:

When a young man kills much meat, he comes to think of himself as a chief or a big man, and he thinks of the rest of us as his servants or his inferiors. We can't accept that. We refuse one who boasts, for someday his pride will make him kill somebody. So we always speak as if his meat is worthless. In this way we cool his heart and make him gentle.²³

As Boehm conceives of it, the egalitarian ethos constitutes a "reverse dominance hierarchy"—the rest of the group acts to subordinate would-be alpha bullies.²⁴ Those who cannot control their dominating tendencies are subjected to a scale of increasing sanctions, from criticism, to ridicule, to ignoring their "orders." And if that is not enough to control would-be bosses,

[o]stracism (taken in a restricted sense as the silent treatment) is one way of putting a deviant on notice, and at the same time of gaining enough distance so that others can be insulated from the aberrant behaviors. ... mild ostracism can allow a political upstart to stay with the group, hopefully to experience some behavioral modifications and gain social reentry. Permanent expulsion from the group, or the group's quietly moving away, carries the distancing still further and suggests that redemption possibilities have been set aside.²⁵

And, as a last resort, would-be bullies have been executed by either the entire group, or selected members.²⁶

It is important that we do not over-stress the egalitarianism of hunter-gatherers. When it comes to food sharing, there is also considerable evidence that in some groups able hunters are often able to control additional resources, and that, generally, high contributors receive additional benefits.²⁷ Like humans everywhere, we find

²³ Lee, *The !Kung San*, pp. 244–6.

²⁴ Boehm, *Hierarchy in the Forest*, p. 87. See also Gintis, van Schaik, and Boehm, "Zoon Politikon: The Evolutionary Origins of Human Political Systems."

²⁵ Boehm, *Hierarchy in the Forest*, p. 77.

²⁶ For data on the frequency of various forms of sanctioning, see Boehm, *Moral Origins*, p. 198.

²⁷ See Gurven, "To Give and to Give Not"; Gurven et al., "Reservation Food Sharing among the Ache of Paraguay," pp. 291ff.

variations. Nevertheless, compared to other types of human social groups, they are remarkably egalitarian social orders.

Boehm's reverse dominance thesis warns us against over-romanticizing small-group egalitarianism. The attempt to constrain domination leads to a general repression of successful individuals. As Jean-Paul Platteau observes in contemporary egalitarian societies:

Compliance with egalitarian norms is not left to the goodwill of the individuals concerned.... [I]t is actually backed by powerful sanctions that comprise social pressures, constant harassment, and the use of effective mechanisms of ideological intimidation. It bears particular emphasis that feelings of envy and jealousy are actually instrumental in bringing about a redistribution of income.... Harming the successful individual is achieved through subtle mechanisms involving witchcraft practices.... In tribal societies, while transient or ordinary luck is believed to result from the normal course of natural events, persisting or exceptional luck is attributed to the obscure and treacherous manipulation of supernatural forces.²⁸

Indeed, "[a] climate of fear is thus created that may prove an awfully effective way of ending deviance."²⁹

7.4. LPA Egalitarianism and Freedom

Critics of the "closed" society such as Hayek and Popper (§1) hold that small-scale egalitarian societies are collectivist, and so are in fundamental tension with the individualism of the Open Society. Ethnographers such as Boehm, Gardner, and Lee dispute this: the egalitarian ethos is not at bottom a "collectivist" ethic of the group subordinating the individual, but one in which the group subordinates those individuals who would control others. As Lee observed:

Egalitarianism is not simply the absence of a headman and other authority figures, but a positive insistence on the essential equality of all people and a refusal to bow to the authority of others, a sentiment expressed in the statement: "... each of us is headman over himself." Leaders do exist, but their influence is subtle and indirect. They never order or make demands of others, and their accumulation of material goods is never more, and often much less, than the average accumulation of the other households in the camp.³⁰

We thus arrive at Boehm's important hypothesis about LPA societies:

[S]uch people are guided by a love of personal freedom. For that reason they manage to make egalitarianism happen, and do so in spite of competitiveness—in spite of

²⁸ Platteau, *Institutions, Social Norms and Economic Development*, p. 201.

²⁹ *Ibid.*, p. 204.

³⁰ Lee, *The !Kung San*, p. 457.

human tendencies to dominance and submission that easily lead to the formation of social dominance hierarchies. People can arrest this process by reacting collectively, often preemptively, to curb individuals who show signs of wanting to dominate their fellows. Their reaction involves fear (of domination), angry defiance, and a *collective* commitment to dominate, which is based on a fear of being individually dominated.³¹

LPA societies appear characterized by a near-obsession with resisting the authority of would-be dominators. Although ethnographers may overstate the case—sometimes downplaying the rigid control over others' lives that characterizes these societies—it in many ways tends to put great stress on preserving personal autonomy.³² "Among foragers and others who are described as pursuing individual autonomy, certain cultural features show up again and again: pressure on children for self-reliance, independence, and individual achievement; individual decision making in matters having to do with family, power, property, ritual, etc.; extreme egalitarianism, including extreme gender egalitarianism; techniques for prestige avoidance and social leveling; absence of leaders."³³ Desmond Jenness, writing in 1922, summed up the views of the Alaskan Eskimos thus: "Every man in his eyes has the same rights and the same privileges as every other man in the community. One may be a better hunter, or a more skillful dancer, or have greater control over the spiritual world, but this does not make him more than one member of a group in which all are theoretically free and equal."³⁴

Yet, again, it is easy for anthropologists to get carried away. The love of freedom and autonomy is achieved through a reverse dominance hierarchy that can bring immense social pressure to bear against anyone who appears to be an "upstart," or "deviant," including those who introduce new techniques and seek ways of improving their own lot. In a highly static environment, this may simply be an effective way for each to protect his own distributive share, but as new opportunities arise (as, for example, in current contexts), it can manifest itself in policing behavior to ensure equality at the cost of improvements for many.³⁵ The extent of gender egalitarianism is also a matter of dispute. Males generally have a privileged role in public and in the family, though female participation in family and political decisions is common. Compared to contemporary America, LPA societies are generally gender inequalitarian; compared to most historical societies, they are often remarkably gender egalitarian. However, we must be careful to keep in mind the distinction between hunter-gatherer bands and the larger clan-based societies that later evolved (§12). These larger societies often engage in sustained warfare, and as warfare comes to dominate, so do males. Thus it is more typical for clan-based societies to stress equality within, but not between, the sexes.³⁶

³¹ Boehm, *Hierarchy in the Forest*, p. 65.

³² Gardner, "Foragers' Pursuit of Individual Autonomy," p. 543.

³³ *Ibid.*, 547–8.

³⁴ Jenness, *The Life of the Copper Eskimos*, p. 94.

³⁵ Platteau, *Institutions, Social Norms and Economic Development*, pp. 206ff.

³⁶ Writing of the Enga, a contemporary clan-based society enmeshed in conflict, Wiessner writes, "Women devote themselves primarily [to] child-rearing, gardening, and pig husbandry, though they are also active in maintaining intergroup ties. The concept of equality of individuals (within sexes, but not between them) and parallel social groups is fiercely defended." "From Spears to M-16s," p. 170.

It should not be surprising that most simple narratives about human egalitarianism are misleading—we are not a simple species. When we understand the Egalitarian Revolution against its distant origins in *Pan*, we see how implausible is the enduring “tribal collectivist” story that we are simply, at bottom, natural egalitarian collectivists. We did not, as it were, commence our life as bees or ants working for the good of the group, but as primates with a keen sense of self-interest, some of whom wish to dominate while others are intent on avoiding their dominance. Ethnographic accounts indicate that hunter-gatherers are devoted to their autonomy and that they have individual lives to lead. Egalitarianism was a method for securing this against threats by would-be dominators. Yet ethnographers tell their own simple story about Egalitarian Revolution, featuring a despotic species that managed to control bullying alphas in the interests of freedom and autonomy in a golden age for humanity. But this was achieved through a reverse dominance hierarchy that could be extremely harsh on deviants who excelled in production or engaged in innovations that upset equal shares. In the current world, such norms can lock groups into poverty.³⁷

§8 Self-Interest, Reciprocity, and Altruism

8.1. Strategies and Evolutionary Outcomes

Up to this point we have been examining the rise of a remarkably egalitarian species out of distinctly non-egalitarian primate. Late Pleistocene egalitarianism was a distinctively human form of cooperation: it was not a brute preference for equal outcomes but a system of cooperative social relations.¹ Underlying the analysis of the rise of egalitarianism is thus a deeper issue: What makes us so cooperative? How did a Machiavellian primate—one motivated to a great extent by self-interest—become a super cooperator? Is egalitarian cooperation really a form of enlightened self-interest? Or have *Homo sapiens* developed new, underlying cooperative motivations? Even more than understanding our egalitarian nature, answering these questions is fundamental to coming to grips with Hayek’s first unsettling thesis: Is our evolved cooperative nature at odds with the Open Society?

The evolution of altruism, reciprocity, and cooperation out of self-interest has been perhaps the dominant foci in human social evolution studies over the last thirty years. Unfortunately, the literature tends to run together strategies for reciprocity and the evolution of reciprocity. While these are sometimes related, at other times they must be kept distinct. The problem is that some cooperative outcome *C* can be said to “evolve” through some strategy *S*—where this can either mean that repeated iterations of *S* produce an unintended social outcome *C*, or that *S* is to be understood as an

³⁷ Platteau, *Institutions, Social Norms and Economic Development*, pp. 206ff.

¹ Gintis, van Schaik, and Boehm, “Zoon Politikon.”

evolutionary mechanism that selects a new basis for cooperative behavior. We must always distinguish:

- I. *The Strategy Claim*: On the basis of genetic-based behavioral trait α under environmental condition E in generation X , strategy S can yield a social outcome C in generation X , still based in genetic-based behavioral trait α .
- II. *The Evolutionary Claim*: On the basis of genetic-based behavioral trait α under environmental condition E in generation X , a population pursuing strategy S can select genetic-based cooperative behavioral trait β in generation $X + n$, where β is distinct from α , and improves cooperation. Cooperation eventually becomes based on β .

Consider, for example, the relation between a tit-for-tat strategy and what Robert Trivers called “reciprocal altruism.”² Both begin at the same place—the Prisoner’s Dilemma (PD), as in Display I.1.

		Player B	
		Cooperate	Defect
Player A	Cooperate	R R	T S
	Defect	S T	P P

where $T > R > P > S$ and $R > (S + T)/2$

Display I.1. Axelrod’s Prisoner’s Dilemma.

As we all know, this is a relentlessly competitive game, in which the best outcome of each is to take advantage (T) of the other, who ends up a sucker (S): in a one-play game the only outcome in equilibrium is “Defect/Defect.” However, as Robert Axelrod, famously demonstrated, in repeated interactions—where players confront each other an indeterminate number of times—cooperation can arise and be sustained via a tit-for-tat strategy, according to which Player A first cooperates, and from then on responds on the next move with whatever play (defect or cooperate) that player B made on the previous move. Later, Ken Binmore stressed the abundance of possible cooperative equilibrium strategies in iterated PDs—indeed, any contract to which rational players might agree to extricate themselves from the “Defect/Defect” outcome is a possible equilibrium strategy in an indefinitely repeated PD.³ Here tit-for-tat is a *strategy* for securing a cooperative social outcome (Cooperate/Cooperate) on the basis of a selfish behavioral trait (thus it is an example of claim I): the players have the same underlying motivations in the n th move as they did on the first.

² Compare Axelrod, “The Emergence of Cooperation among Egoists,” and Trivers, “Evolution of Reciprocal Altruism.” See also Nowak and Highfield, *Super Cooperators*, chap. 1; Cosmides and Tooby, “Cognitive Adaptations for Social Exchange,” pp. 170–9.

³ Binmore, *Natural Justice*, pp. 79–82.

Trivers based his evolutionary account on essentially the same game (though he often supposed it was a sequential game, in which one player moved first).⁴ His examples included food sharing in humans, groupers and cleaning fish, and warning calls in birds. Suppose A shares food with B at time t_1 ; this increases the fitness of B, but A gives up some of his food. It is rational/fitness enhancing for “selfish” A to so act only if B returns the favor at time t_2 ; otherwise A gets the “sucker” S payoff. But why would B refrain from suckering A? Clearly, if doing so secures yet more (future) assistance from A. We can also think of this in terms of punishment: A can essentially “punish” the lack of reciprocity (defection) from B at t_2 by withholding future assistance at t_3 .⁵ And, critically, the fitness of both is increased by the $\{R, R\}$ outcome over the $\{P, P\}$ outcome,⁶ so both are better off by assisting each other than by mutually defecting. Thus the Evolutionary Claim: a new underlying behavioral trait of “helping the other” could arise in subsequent generations, as those with the helping trait displaced those without it.

It can be easy to confuse these analyses—indeed, both are referred to under the moniker “evolution of cooperation.” Suppose that we discover a group of people who share food in a “direct reciprocity” pattern.⁷ It is tempting to see this as significant evidence for the Evolutionary Claim (II), but it is at least as strong evidence for the Strategy Claim (I). If A gives to B in a way that closely tracks what B gave to A, this may tell us little about an evolved behavioral trait, but a great deal about the dyadic relations and strategies between agents in the current group. On the Strategy Claim, if those dyadic relations change, the helping behavior would also quickly change. The Evolutionary Claim must show that some new trait—e.g., some behavioral tendency that does not reduce to self-interest under a set of norms—has developed via the mechanism. In that case, we would expect that the new behavioral trait may persist long (perhaps indefinitely) after the underlying reciprocity relations have changed.

8.2. The Invisible Hand?

The distinction between these two claims is of the first importance. Understood as an instance of the Strategy Claim, analyses of human reciprocity explain how, via intentional strategies, norms, or institutions (such as the state), agents who are still to a large extent Machiavellian egoists can nevertheless engage in ultra-social cooperation. For many economists, the invisible hand of the market is just such a mechanism: “Even in the world of *knaveish* men, Smith demonstrated that economic liberalism enabled peaceful social cooperation that leads to increases in productivity. Indeed, he pointed out that *liberalism could not only deal with a world of selfish individuals, but actually harnessed man’s self-interested motivation for the benefit of everyone*. Under liberalism, *self-interested and*

⁴ Display I.1 does not capture the sequential character of Triver’s game. For present purposes, we can set aside this formality.

⁵ As Binmore has stressed, this is an application of the “folk theorem.” See *Natural Justice*, pp. 79–82.

⁶ $\{R, R\}$ is also better than taking turns playing each other for a sucker, as specified by Axelrod’s $(R > (S+T)/2)$.

⁷ See, for example, Gurven et al., “Reservation Food Sharing among the Ache of Paraguay.”

rapacious man is ‘led by an invisible hand to promote an end that was no part of his intention’—to the interest of society.”⁸

If this were so, we would not need to suppose any evolved change in basic motivational structure to explain current human ultra-sociality, much less Pleistocene cooperation. On this view, what has evolved are institutions and strategies for channeling the activities of a basically self-interested primate, not a radical change in the cooperative motivations themselves. However, though it is without doubt true that direct reciprocity mechanisms characterize both hunter-gatherer⁹ and contemporary societies,¹⁰ it is doubtful that they can sustain ultra-social cooperation. There are a number of reasonably demanding conditions necessary to generate cooperation based on direct reciprocity. There must be an extended ongoing series of interactions, so A and B must not be too dispersed. Moreover, the form of cooperation must be resistant to cheating strategies such as “shaving” reciprocated benefits: if the degree of received benefit is highly continuous or hard to ascertain, it is in B’s interest to return slightly less; reciprocated shaving is apt to undermine long-term reciprocation.¹¹ In addition, direct reciprocity is difficult to scale up to even modest-sized groups, being basically about dyadic interactions;¹² it requires extensive memory and record keeping of past interactions with all partners. To be effective, there must be an extended series of interactions in which comparable benefits at comparable costs are exchanged.

8.3. Evolved Strong Reciprocity

In recent years, an impressive body of both experimental data and evolutionary models has provided evidence that humans are *Strong Reciprocators*.¹³ On this view, humans *now*, to a significant extent, tend to respond to cooperation with cooperation and tend to inflict punishment on those who cheat on cooperative rules and practices. In this sense, we have a marked tendency to be conditional cooperators: one cooperates if others do too. Between *Pan* and, say, 100,000 years ago, we became less Machiavellian self-interested agents: underlying traits changed, as per the Evolutionary Claim, II. To be sure, this reciprocity is overlaid on the more basic egoism of *Pan*. “Humans have evolved a social psychology that mixes a strong element of cooperative dispositions ... with an equally strong selfish element deriving from our more ancient primate dispositions.”¹⁴ While, of course, we are often selfish (and a few of us are almost

⁸ Boettke, *F. A. Hayek*, p. 134. Emphasis added.

⁹ Gurven et al., “Reservation Food Sharing among the Ache of Paraguay.”

¹⁰ Henrich and Henrich, *Why Humans Cooperate*, pp. 48–58, 116–23.

¹¹ A good (and literal) example is the shaving or clipping of money when coins were essentially a certified weight of precious metals. Traders shaved off an imperceptible amount of silver or gold, thus constantly reducing the real payment constituting, as Locke said, a sort of robbery, not reciprocity. Locke, *Some Considerations*, p. 75.

¹² See Bowles and Gintis, *A Cooperative Species*, pp. 63–8; Henrich and Henrich, *Why Humans Cooperate*, p. 51.

¹³ For useful overviews, see chapters 5–8 in *Moral Sentiments and Material Interests: The Foundation of Cooperation in Economics Life*, edited by Herbert Gintis, Samuel Boyd, and Ernst Fehr. See also my *Order of Public Reason*, pp. 101–22, and the accompanying citations. For a somewhat dissenting analysis, see Tomasello, *A Natural History of Morality*, chap. 2. In this subsection I consider the “reciprocator” part of the description; the “strong” part is deferred until section 9.1.

¹⁴ Richerson and Boyd, “The Evolution of Free Enterprise Values,” p. 114. To conjecture on evolutionary mechanisms: it is plausible that the extraordinarily harsh Pleistocene era strongly selected for more

always selfish), most are often willing to cooperate *when* others cooperate. Thus we add another layer to our basic cooperative nature: by the Late Pleistocene period we were not only egalitarians, but strong reciprocators.

Richerson and Boyd see conditional cooperation as basic to the “moral hidden hand”¹⁵ that allows the market to function. *Pace* the traditional reading in economics, it is also at the heart of Smith’s analysis of the market. For Smith, it is the deep and unique human tendency—perhaps “an original” principle of human nature—to gain through *exchange* that is at the core of markets and the division of labor.¹⁶ Smith stresses that it is our tendency to pursue our interests through exchange, not our tendency to maximize self-interest, that drives economic life and encourages the growth of extensive markets. He is explicit that individuals, specializing through the division of labor, need each other’s cooperation,¹⁷ and so in trading are playing a cooperative game.¹⁸ As Brian Skyrms and his students have emphasized, such reciprocal cooperators tend not to play Prisoner’s Dilemmas but more often Stag Hunts, as in Display I.2.¹⁹

		Player B	
		Stag	Hare
Player A	Stag	1 st 2 nd	3 rd 2 nd
	Hare	3 rd 2 nd	2 nd 2 nd

Display I.2. A Stag Hunt.

cooperative agents and groups, intensifying the already present trend to more cooperative motivations. See Bowles and Gintis, *A Cooperative Species*, chap. 6; Richerson and Boyd, *Not by Genes Alone*, pp. 224–9; Richerson and Boyd, “Rethinking Paleoanthropology: A World Queerer than We Supposed”; Boehm, *Moral Origins*, chap. 10.

¹⁵ Richerson and Boyd, “The Evolution of Free Enterprise Values,” p. 116.
¹⁶ Smith, *Wealth of Nations*, vol. I, p. 17.
¹⁷ *Ibid.*, vol. 1, p. 18.
¹⁸ To be sure, we must not confuse reciprocity with altruism. Much of the evolution of cooperation literature has focused on the idea of evolution of altruism—one agent sacrificing some of her fitness for another. This can refer to a wide variety of cooperative behaviors, from helping kin, or partners in joint activities, to unrequited self-sacrificing behavior. Reciprocity or conditional cooperation can be viewed as a (moderate) type of altruism (as is any departure from pure egoism). But reciprocity is certainly distinct from the psychological state of feeling altruistic or benevolent toward others or a tendency to engage in unrequited helping. Although humans are manifestly capable of such feelings and actions, Smith was quite right to stress that they are not the foundation of the division of labor and social cooperation: “It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest.” (*ibid.*, vol. 1, p. 18) This passage is often mistakenly read as claiming that pure self-interest is enough (as is claimed in the selfish version of the invisible hand). Once we clearly distinguish egoism, reciprocity, and benevolence, we can see that denying that benevolence moves us in our normal interpersonal relations is not to embrace the primacy of egoism. The core idea of reciprocity is clear in the sentence immediately preceding: “Give me that which I want, and you shall have this which you want . . .” That is not something that players in a one-play Prisoner’s Dilemma can honestly promise. McCloskey’s analysis identifies the error of identifying economic agents as resolutely selfish, though she does rather rush to virtue and altruism as the alternative. *Bourgeois Equality*, pp. 336–44. See also my “The Open Society as a Rule-Based Order.”
¹⁹ Skyrms, *The Stag Hunt and the Evolution of Social Structure*; Alexander, *The Structural Evolution of Morality*, pp. 102ff., 238ff.

In this game, both players' first choice is to hunt stag together: mutual cooperation can bring higher gains than solo hunting of hare, but it takes two to successfully hunt the stag (a division of labor is required). But to do one's part in the stag hunting effort when the other does not is the worst option: one has wasted one's effort on an unsuccessful attempt to cooperate. Thus there are two equilibria in this game: the "payoff dominant" Stag/Stag and the "risk dominant" Hare/Hare. The problem here is not that each is tempted to defect, but that each needs to trust the other to do their part if both are to secure the payoff dominant equilibrium.

The rational play in this game is by no means trivial: it is all too easy for a population to spend most of its time hunting hare.²⁰ It is critical that individuals successfully signal their trustworthiness to each other.²¹ But, as Smith was the first to stress, among conditional cooperators intense social cooperation through the division of labor is not at all mysterious. Once cooperation is established it tends to stably proceed, to expand, and to benefit all. Each individual is constantly adapting to the ongoing division of labor, and the mutually beneficial exchanges on offer.

On Boehm's analysis, selection for reciprocity was another factor at the heart of Pleistocene social evolution. This is often called "indirect reciprocity:" one's status and reproductive opportunities are influenced by one's general reputation for being the sort of person who reciprocates.²² An especially interesting study of the social consequences of having a reputation for contributing to others' welfare was conducted among the Ache of Paraguay.²³ The Ache were full time hunter-gatherers until the 1970s; they now live in permanent settlements, but still spend about a third of their time hunting. Michael Gurven and his coworkers studied aid by the group to those who were injured or ill. Various forms of injury and illness are common among hunters, so the informal social insurance of food aid by others is critical to medium-term well-being. Gurven and his coworkers divided the Ache into four groups: (1) the *Philanthropic*, who are both high producers and have been known to give a large proportion of their food to others in need; (2) the *Greedy*, who produce much but give a small proportion of their food; (3) the *Means-well*, who produce little but give a high proportion of what they do produce; and (4) the *Ne'er-do-wells*, who neither produce nor give much. Although the *Greedy* give a relatively small share of their food to others, this is still an absolutely larger amount than the *Means-well* give. It was found in this study that, when in need, the *Philanthropic* and *Means-well* well group received more help when needed than the *Greedy* or *Ne'er-do-wells*. However, the *Philanthropic*, even though valuable high producers, did not receive more help than the *Means-well*; and *Ne'er-do-wells* consistently received more help than *Greedy*, even though in absolute terms the *Greedy* gave more food to others than did the *Ne'er-do-wells*. At least when it comes to aid to those in need, what seemed most important is the readiness

²⁰ Skyrms, *The Stag Hunt and the Evolution of Social Structure*, chap. 3.

²¹ Ibid., Part II. See also Skyrms, *Signals*. Public rituals may be helpful in providing common knowledge that all wish to cooperate. See Chwe, *Rational Ritual*, esp. pp. 25ff.

²² Boehm, *Moral Origins*, chap. 7; Alexander, *Human Social Evolution*, chap. 9; Henrich and Henrich, *Why Humans Cooperate*, pp. 58–64; Nowak and Sigmund, "Evolution of Indirect Reciprocity."

²³ Gurven et al., "'It's a Wonderful Life': Signaling Generosity among the Ache of Paraguay." See also Boehm, *Moral Origins*, chap. 11.

of the needy to help others; those well off who were skimpy in helping were the least likely to be assisted. Those known to be selfish are thus at risk. At least in this case of “social insurance,” help is not attuned to actual benefits received in the past (as the direct reciprocity strategy would indicate), but rather to the party’s reputation as someone willing to assist others. Overall, there is widespread agreement that one of the basic postulates of egalitarian society is the social importance of generosity and reciprocity.²⁴

It is critical to stress that this is not a *strategy* to secure reciprocal benefits, but an evolved social disposition. In the philosophic literature on morality and cooperation, a common objection is that a “mutual benefit” analysis of morality cannot account for moral obligations to those unable to cooperate. Gauthier, who adopts a strategy-based analysis of cooperation (Claim I as noted earlier), seems to accept this implication.²⁵ Whatever the merits of this objection to a mechanism (or strategy) grounded analysis of cooperation, it does not apply to an evolved reciprocity analysis. For a wide variety of reasons, the evolution of reciprocity need not (and apparently often does not) include a disposition to track individual contributions: what may well be critical is that the group’s social life is widely informed by a reputation for reciprocity and assisting others in need. Thus, as with the Ache, those who display a tendency to assist others may well be preferred over high contributors.

§9 Internalized, Enforced, Social Rules

9.1. The Case of Cephu

Of course in any social system—certainly including LPA societies—people sometimes find themselves with opportunities to cheat. One can get the benefits without paying the costs; as I have been stressing, human reciprocity is, as it were, built on a Machiavellian chassis. When the potential gains are high, many are tempted to cheat, and some are always on the lookout for such opportunities. Strong reciprocators are not only conditional cooperators, but are ready to expend some of their own resources in punishing those who defect from cooperative arrangements.¹ As I have stressed elsewhere, the evolution of agent types who are prepared to punish to enforce the basic rules of cooperation is a fundamental feature of human social life.² Given this, it is tempting to think that the effective threat of punishment suffices to overcome the limitations of conditional cooperation. If we look simply at obvious costs and benefits,

²⁴ Flannery and Marcus, *The Creation of Inequality*, pp. 33, 74.

²⁵ Gauthier, *Morals by Agreement*, p. 268. Peter Vanderschraaf has devised a compelling reply to this “silver bullet” to justice as mutual advantage. *Strategic Justice*, pp. 280ff.

¹ See, for example, Boyd, Gintis, Bowles, and Richerson, “The Evolution of Altruistic Punishment.” See also my “Retributive Justice and Social Cooperation.”

² *The Order of Public Reason*, pp. 118–22.

sufficient punishment can sustain any type of order.³ If the costs of punishment are sufficiently high, it looks obvious that individuals will obey just about any social rule. This would seem especially so in small LPA societies. In such societies, everyone's actions are generally known to all, and so all can be policed by all. Such was Russell Hardin's view of small group social life:

An individual need not rely on self-regulation to be moral. The knowledge that the whole community has of each individual's adherence to the local moral code allows community members to sanction miscreants. An enormous part of the debate about morality in the modern secular world is about how individuals can be motivated to act morally. That question is answered easily [in small-scale society]... *The community spontaneously enforces its morality as a set of compulsory norms.*... The exaction would typically be quick and aimed at the right person.⁴

I believe this common view is mistaken. Punishment is disruptive and dangerous to small groups such as LPA societies. It can lead to feuds or dissolution of the group—and given that groups are often on the margin of survival, a breakup can mean the death of all.⁵ The decision to punish is thus not one that the group spontaneously and easily makes, but a deliberate and essentially political process that requires formation of a group consensus.⁶

These political dynamics are striking in Colin Turnbull's famous case of Cephu, the cheating hunter. The Pygmy hunters studied by Turnbull sometimes hunt small game with nets. The men place their nets in a long semi-circle, and women and children drive game into the nets. Cephu, having complained of consistent bad luck in hunting, decided to secretly put his nets in front of the others, so game would be first driven into his net. This worked in increasing his catch but, unfortunately for him, he was observed. Turnbull continues the account as the hunters

strode into camp with glowering faces and threw their nets on the ground outside their huts. Then they sat down, with their chins in their hands, staring into space and saying nothing. The women followed, mostly with empty baskets, but they were by no means silent. They swore at each other, they swore at their husbands, and most of all they swore at Cephu.

I tried to find out what had happened, but nobody would say. Kenge, who had been sleeping, came out of our hut and joined the shouting. He was the only male who was not sitting down, and although he was young he had a powerful voice, and a colorful use of language. I heard him saying, "Cephu is an impotent old fool. No, he isn't, he is an impotent old animal—we have treated him like a man for long enough, now we should treat him like an animal. Animal!" He shouted the final epithet across at Cephu's camp, although Cephu had not yet returned.

³ Boyd and Richerson, "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups."

⁴ Hardin, "The Priority of Social Order," p. 412. Emphasis added.

⁵ See Nikiforakis, Noussair, and Wilkenin, "Normative Conflict and Feuds: The Limits of Self-Enforcement."

⁶ See Boehm, *Hierarchy in the Forest*, p. 118.

The result of Kenge's tirade was that everyone calmed down and began criticizing Cephu a little less heatedly, but on every possible score: The way he always built his camp separately, the way he had even referred to it as a separate camp, the way he mistreated his relatives, his general deceitfulness, the dirtiness of his camp, and even his own personal habits. . . .

...

Trying not to walk too quickly, yet afraid to dawdle too deliberately, he [Cephu] made an awkward entrance. For as good an actor as Cephu it was surprising. By the time he got to the *kumamolimo* everyone was doing something to occupy himself—staring into the fire or up at the treetops, roasting plantains, smoking, or whittling away at arrow shafts. Only Ekianga and Manyalibo looked impatient, but they said nothing. Cephu walked into the group, and still nobody spoke. He went up to where a youth was sitting in a chair. Usually he would have been offered a seat without his having to ask, and now he did not dare ask, and the youth continued to sit there in as nonchalant a manner as he could muster. Cephu went to another chair where Amabosu was sitting. He shook it violently when Amabosu ignored him, at which he was told, "Animals lie on the ground."

...

Cephu tried very weakly to say that he had lost touch with the others and was still waiting when he heard the beating begin. It was only then that he had set up his net, where he was. Knowing that nobody believed him, he added that in any case he felt he deserved a better place in the line of nets. After all, was he not an important man, a chief, in fact, of his own band.

Turnbull goes on:

Cephu knew he was defeated and humiliated. Alone, his band of four or five families was too small to make an efficient hunting unit. He apologized profusely, reiterated that he really did not know he had set up his net in front of the others, and said that in any case he would hand over all the meat. This settled the matter, and accompanied by most of the group he returned to his little camp and brusquely ordered his wife to hand over the spoils. She had little chance to refuse, as hands were already reaching into her basket and under the leaves of the roof where she had hidden some liver in anticipation of just such a contingency. Even her cooking pot was emptied. Then each of the other huts was searched and all the meat taken. Cephu's family protested loudly and Cephu tried hard to cry, but this time it was forced and everyone laughed at him. He clutched his stomach and said he would die; die because he was hungry and his brothers had taken away all his food; die because he was not respected.

... From Cephu's camp came the sound of the old man, still trying hard to cry, moaning about his unfortunate situation, making noises that were meant to indicate hunger. From our own camp came the jeers of women, ridiculing him and imitating his moans.⁷

⁷ Turnbull, *The Forest People*, pp. 104–8.

Note first that the group decides whether a violation has occurred. Often the lead is taken by one individual, in this case Kenge, who is not necessarily the directly injured party. This helps ensure that the dispute will not simply be seen a dyadic conflict. Consensus forms that a violation has occurred; note especially that while Cephu's family does not join in the punishment, neither do they resist. Because small-scale societies are a complex mix of kin and non-kin relations, it is important that punishment does not lead to inter-family conflict. This is especially clear in cases of capital punishment, which is practiced in many hunter-gatherer societies.⁸ In cases of capital punishment, the entire group of males, including the victim's kin, sometimes collectively kills the offender (in one noted case, the entire group, including women, participated in the execution). In many cases a kin of the offender is selected as executioner.⁹ The critical point here is that because eruption of counter-sanctioning is always a possibility, the rule enforced must be seen by all as legitimate, it must be agreed that a violation has occurred, and the kin of the deviant must at least passively accept, and sometimes must actively participate in, the punishment. Lethal weapons abound in hunter-gatherer groups, and the escalation of violence is an ever-present threat.

As Bowles and Gintis more generally stress, effective punishment depends on legitimacy: unless those to be punished and their friends and allies are convinced that the rule being enforced is a legitimate one and one for which community enforcement is appropriate, a punishing action taken as a means to protect social cooperation can lead to weakening it.¹⁰ Experimental evidence confirms that attempts at punishment readily evoke counter-punishment when the offender does not experience guilt.¹¹

9.2. The Internalization of Moral Rules

Note that with Cephu the admission of guilt preceded the group's confiscation of his kill. Consensus on the lower levels of punishment, ridicule, and mild ostracism were reached during the walk home and afterward, and it is this less dangerous level of punishment that triggered his profuse apologies—and only after that did confiscation occur.

The important point is how costly such punishing episodes are to the group. Hunting is an egalitarian, cooperative activity, and shirkers, cheats, and free-riders such as Cephu pose real threats. Cephu, indeed, not only posed a threat by cheating, but he initially resisted punishment and sought to intimidate others, arguing that he was an important person, indeed a chief.¹² Rules that are generally perceived as purely external—simply a social fact about when one is liable to punishment—by group

⁸ Boehm reports that in his database about half the hunter-gatherer societies coded practice capital punishment; there is strong reason to think that the number may be much higher, as central governments treat band and tribal executions as murder. *Moral Origins*, p. 84.

⁹ Boehm, *Hierarchy in the Forest*, pp. 81–2, 121–2, 180. While females seldom participate in the executions, they do typically participate in the deliberation leading to execution.

¹⁰ Bowles and Gintis, *A Cooperative Species*, p. 26. As Bowles and Gintis point out, in large-scale societies, too, anti-social punishment (counter-punishment) is real: experiments show great differences in societies in the extent to which punishment is accepted or evokes counter-sanctioning.

¹¹ See Hopfensitz and Reuben, "The Importance of Emotions for the Effectiveness of Social Punishment"; Bicchieri, Dimanta, and Xiao, "Deviant or Wrong?"; Nikiforakis, Noussair, and Wilkening, "Normative Conflict and Feuds."

¹² Boehm, *Moral Origins*, p. 43.

members, depending solely on self-interest to motivate compliance, would be a hopelessly inefficient way of securing cooperation, inviting both opportunistic evasion and counter-punishment. The large majority must, and do, internalize the rules, which involves emotional attachment to the rules and compliance with them.¹³ As we saw, the hierarchical nature of *Pan's* (§6.3) social life probably provided a pre-adaptation for the learning of, and conformity with, cooperative social rules. Individuals who conform have a virtue highly prized in many small hunter groups—self-control.¹⁴ In the face of temptations to cheat and dominate, they can be counted on to generally comply with the group's rules. Cephu was lacking in self-control and was a severe problem for the group: he needed watching. Those even more seriously lacking in self-control, such as repeated murderers, can be executed.¹⁵ Overall, hunter-gatherer societies display a high level of rule internalization and corresponding self-control.

As Darwin suggested, a definitive development in the moral sense of humans was the development of conscience or, more accurately, internalized normative guidance.¹⁶ But this internal guidance is rule-based: it is an ability to follow the social rules of our group. "This view suggests that, stripped of our social norms and beliefs, we aren't nearly as cooperative or as communal as we might seem."¹⁷ Individuals not only see the rules of morality as external guidelines as to how they are expected to behave, but also adopt the guidelines as internal demands they make upon themselves, and feel guilt and shame when they fail to conform. Indeed, unless a creature can regulate his behavior through internalized prescriptions addressed to herself, it is doubtful that we would say that she is a moral agent.¹⁸

Students of cognition have recently turned to modeling the processes that underlie norm internalization.¹⁹ We know that internalization of moral rules is a normal accomplishment for humans, and occurs at a very young age. In a series of experiments conducted by Nunnar-Winkler and Sodian, children between four and eight years old were told a story about two children, both of whom liked candy. The first child was tempted to steal the candy, but did not; the second stole the candy. Even the four-year-old subjects knew that stealing was wrong and could provide reasons why this is so. Thus, they could engage in punishing violators. The difference is that the youngest children expected the child who stole the candy to be happy with his violation of the rule, while they (the youngest children) expected the child who resisted temptation to be sad. Older children reversed this; they supposed the child who stole would be sad—guilty—while the child who resisted temptation would be the happy one. Younger children apparently expect people to be happy when they get what, all things considered, they want, regardless of whether this violates a moral requirement and harms others.²⁰ Again, older children

¹³ Ibid., pp. 113–4.

¹⁴ For a striking case, see Boehm, *Hierarchy in the Forest*, pp. 51–9.

¹⁵ In Boehm's database, of the societies that engaged in capital punishment, repeat murder was the second most reported capital offense.

¹⁶ Darwin, *The Descent of Man*, chap. V. See also Kitcher, *The Ethical Project*, chap. 2; Boehm, *Moral Origins*, chaps. 1 and 2.

¹⁷ Henrich, *The Secret of Our Success*, p. 154.

¹⁸ See Joyce, *The Evolution of Morality*, pp. 101–5.

¹⁹ See Andrighetto, Villatoro and Conte, "Norm Internalization in Artificial Societies."

²⁰ It is generally thought that young children see harm to others as violating a basic moral requirement. See Turiel, Killen, and Helwig, "Morality: Its Structure, Functions and Vagaries," p. 174. Guilt is especially associated with violation of rules against harm and the rights of others. Jesse Prinz, *The Emotional Construction of Morals*, p. 77. See §18.4 in this volume.

expected the violator to feel unhappy. Nunnar-Winkler and Sodian conclude, “children may first come to know moral rules in a purely informational sense, that is, they know that norms exist and why they should exist. Not until several years later, however, do they seem to treat them as personally binding obligations the intentional violation of which will be followed by negatively-charged self-evaluative emotions or genuinely empathetic concerns.”²¹

Very young children view moral rules as simply external guides. They can appreciate reasons that these rules are important and even that punishment is appropriate; what they do not grasp is that the rule can function as a requirement in an agent’s deliberations and can be seen as “personally binding,”²² so that the agent will feel guilt for failing to meet this requirement even if by so doing she gets what she wants. What very young children do not grasp is that a typical moral agent cares about moral requirements and so can put aside the things that she wants and, instead, conform to the rule’s requirements, and success in doing this relates to her own self-esteem. As Abraham Lincoln was said to have remarked, “when I do good, I feel good. When I do bad, I feel bad. That is my religion.”²³

9.3. The Necessity of Reconciliation

We are beginning to see a much more complicated picture of small-group egalitarian morality.²⁴ As modern humans emerged 40,000 years ago, they were a complex admixture of self-interested, strategic, Machiavellian, and conditional cooperators intent on avoiding domination. They were ready to live together and generally do their part of shared cooperative rules that they internalized, yet because they were also self-interested they were on the lookout for opportunities to cheat. As Boehm nicely puts it, we have “flexible consciences.”²⁵ More subtly, self-interest infects our judgments about the best social rules—our moral judgments. As philosophers from Hobbes to Rawls have noted, each tends to favor understandings of moral rules that advance her interests,²⁶ a common-sense belief for which there is experimental support.²⁷ So by its very nature, human social life is characterized by the necessity of a shared morality and a tussle over what these common rules will be. But if this tussle goes too far, our social existence is impaired. Consequently, from hunter-gatherers onward, a critical part of social life has been to somehow channel these divergent interests and views into consensual decisions about the rules.²⁸ Normative disagreement is not a modern invention: it is built into our nature as ultra-social beings. Thus, we might say, from

²¹ Nunner-Winkler and Sodian, “Children’s Understanding of Moral Emotions,” p. 1336. Emphasis in original.

²² *Ibid.*, p. 1324.

²³ See Bowles and Gintis, *The Cooperative Species*, p. 169. Bowles and Gintis devote great care to analyzing how internalization of social morality can be modeled (chap. 10).

²⁴ Parts of this section draw on “Laws, Norms, and Public Justification,” written with Jacob Barrett.

²⁵ Boehm, *Moral Origins*, pp. 29ff.

²⁶ Hobbes, *Leviathan*, p. 180; Rawls, *A Theory of Justice*, pp. 171–2, 195–6.

²⁷ See, e.g., Bicchieri and Chavez, “Norm Manipulation, Norm Evasion”; DeScioli, “The Side-Taking Hypothesis for Moral Judgment,” p. 25.

²⁸ Gintis, van Schaik, and Boehm, “*Zoon Politikon*.”

the very first, human morality has been a reconciliation project. Early human society was not a submersion of individual personality into a homogeneous group.

To some extent, humans living under cooperative rules must reconcile: it is not enough for some to simply declare what the rules will be: they must be widely embraced. A critical function of moral rules is to promote two types of shared expectations: (i) shared *empirical expectations* about what, in a some circumstance *C*, others *will* do and (ii) shared *normative expectations* about what in *C* others think I must do.²⁹ When we share such rules, we have reconciled our different views about what the rules should be, and so can we coordinate our actions and moral claims. For such a system to effectively function, not only must individuals share these expectations, but they must, in addition, be *sensitive* to the demands of the rules in the sense that they are willing to forgo personal gains to comply with the shared rules. Sensitivity is so important because, as ambivalent cooperators with flexible consciences, we are often pulled between rule compliance and opportunistic cheating. To be effective as a basis for social cooperation, individuals must generally be sensitive to the rule: not only must they internalize it, but they must be willing to forgo self-interested gains.

Bicchieri has developed a sensitivity variable to measure this: How much material gain is the person willing to forgo for the sake of acting on a norm? On her influential analysis of a social norm (or, alternatively, a social-moral rule),³⁰ a person whose personal normative convictions support a norm will be more sensitive to its requirements: he will be willing to pay a greater personal cost in order to adhere to it, and so will be more likely to comply with it, even in the absence of the threat of sanctions.³¹ Bicchieri explains:

Sensitivity to a norm refers to how much a person adheres to what the norm stands for. Norm sensitivity embodies one's personal reasons for adhering to the norm. A highly sensitive individual could list several good, important, reasons why a particular norm should be enforced, whereas an individual with low sensitivity, who does not care much about what the norm stands for, may only list the fact that, since the norm is widespread, it makes sense for her to obey it (to avoid the sanctions that transgressions incur). Let us call a person's sensitivity to a particular norm, n , k_n . For example, a person who is not very convinced of the advisability of child marriage will have very low sensitivity to that norm (in other words a very low k_n), whereas a person who is convinced that child marriage is the best way to protect a child's honor will be highly sensitive to the norm.³²

A person who is sensitive to a social or moral norm is one who believes there are many "good reasons" for adhering to and, presumably, enforcing the norm. A person who is highly sensitive to a norm—who is willing to follow it even at considerable cost to

²⁹ See Gaus, *The Order of Public Reason*, pp. 162ff.

³⁰ For the relation between my understanding of a "social moral rule" and Bicchieri's understanding of a social norm, see *ibid.*, pp. 163–72. For present purposes, they can be treated as essentially equivalent.

³¹ In Bicchieri's formal and empirical work, the sensitivity variable (k) measures a person's tendency to forgo monetary gains in order to comply with a fairness norm. Bicchieri, *The Grammar of Society*, pp. 52–4.

³² Bicchieri, *Norms in the Wild*, p. 165.

herself—is likely to be one for whom it is justified in the sense that the norm “stands for” or promotes the things she cares about.³³ Let us call this:

The Justification Effect: one’s sensitivity (k) to a moral rule/norm tends to rise as its justification increases, where justification depends on the coherence of the rule/norm with one’s own personal normative convictions.

This is only a minimal type of justification: a moral rule is justified to a person if it aligns with her normative convictions. Such justification does not interrogate the grounds of those convictions—whether a person’s normative convictions are, say, themselves based on badly grounded beliefs. Nevertheless, this is a critical insight. *From the very beginning, human morality has relied on “public justification”:* the rules of the group must be such that the members’ “personal normative convictions” and interests align with them. We have already seen that hunter-gatherers are deeply concerned with the freedom and autonomy (§7.4). The idea that freedom and autonomy are Western inventions, or that the concern for publicly justified moral rules was discovered by social contract theory (or, even more implausibly, by John Rawls), fails to appreciate that they have been fundamental features of moral life from the beginning. We must always keep in mind that egalitarian small-scale society evolved from individualistic, strategically minded *Pan*.

Bicchieri also points to a more demanding notion of justification.³⁴ One can come to recognize that one’s personal normative convictions are themselves not well-grounded, for example when we realize that our empirical beliefs and normative commitments do not cohere.³⁵ Bicchieri and Hugo Mercier thus argue:

Inconsistencies are typically the occasion for belief change. When inconsistent beliefs are detected, the mind tries to determine which can be most easily rejected in order to reduce the inconsistency. . . . Arguments take a belief that the listener accepts—the premise—and show her that this belief is inconsistent with the rejection of the argument’s conclusion. When a good argument is offered, it is more consistent for the listener to change her mind about the conclusion than to accept the premise while rejecting the conclusion.³⁶

While unreflectively one may conclude that one’s personal normative beliefs endorse a norm, upon further argumentation or reflection on relevant data, one may come to

³³ One can view these things a person cares about as “commitments.” See Sen, “Rational Fools.” From another perspective, this connects with what Rawls calls the need for “congruence” between the right and the good. *A Theory of Justice*, pp. 450ff.

³⁴ See Bicchieri and Mercier, “Self-Serving Biases and Public Justifications in Trust Games”; Bicchieri and Mercier, “Norms and Beliefs.”

³⁵ Bicchieri, *Norms in the Wild*, pp. 129–30. Those with greater “reflective autonomy,” Bicchieri predicts, will have a stronger tendency to decrease their sensitivity to a norm as they become aware of reasons against it, while more conformist members of the group will have higher sensitivity just because, say, the norm has been in place for a long time, and so will be less sensitive to reasons against it. Bicchieri, *Norms in the Wild*, pp. 166ff. On conformity, see §11.2 in this volume.

³⁶ Bicchieri and Mercier, “Norms and Beliefs,” p. 69.

see either that this is not so, or that one's moral convictions were flawed. When successful this leads to:

Robust Public Justification: a moral rule/norm is robustly justified in a social group *G* if (i) at least a large majority of *G* view their personal normative beliefs as giving reasons to hold that everyone in the group ought to act on the moral rule/norm, and (ii) this conclusion is stable in the light of the amount of reflection on their beliefs, discussion, and exposure to new information that it is reasonable to expect of typical members of *G*.

Obviously clause (ii) is contextual and rather vague.³⁷ The root idea, though, is that in any given case, a moral rule fails to be robustly justified if, in light of the degree of critical reflection and discussion that is appropriate to the group on this matter, they conclude that their personal normative beliefs do not give them grounds to endorse it.³⁸

Robust Public Justification is no mere philosopher's will-o'-the-wisp: it is, essentially, the aim of the Tostan Community Empowerment Program.³⁹ The program, as conducted in rural Senegal (in villages ranging from 200 to 500 people),⁴⁰ centers on human rights and democracy education, stressing the exploration of, and deliberation about, the values recognized by the members of the community. Throughout the curriculum, the aim is to examine these ideas in light of the values of the community members. The participants in these classes reflect on human rights and equality (for example, concerning gender norms), often reaching considerable consensus within the group about these values and some of their implications, before going out to engage in further deliberation and discussion with the wider community. As I have been stressing, disagreement, discussion, and deliberation are neither modern nor Western inventions: they are extensive in small-scale societies, including LPA social orders. As we saw with the case of Cephu, even obvious violations call for group-wide discussions.

As we move from an unjustified moral rule to a minimally justified one, and then on toward Robust Public Justification, not only is the autonomy of the rule follower respected,⁴¹ but the moral rule/norm becomes stronger and more stable in at least three ways. *First*, dissemination of new information is apt to confirm endorsement of the rule: it is not based on insulating error from interrogation. *Second*, since personal normative convictions are firmly aligned with the moral rule, individuals are typically more sensitive to it, and informal (and formal) punishment becomes less important as agents become less tempted to defect. *Third*, critical reflection and discussion are likely to enhance rather than undermine an individual's normative convictions and therefore to enhance individuals' sensitivity to the norm. So robustly justified moral rules are apt to be more stable in the face of the spread of information, temptations to defect, and critical reflection and discussion. Their efficacy and stability are not

³⁷ For some philosophical cleaning up, see Gaus, *The Order of Public Reason*, pp. 254–8.

³⁸ See Cislighi, Diane, and Mackie, *Values Deliberation and Collective Action in Rural Senegal*; Bicchieri, *Norms in the Wild*, pp. 132ff, 159–69.

³⁹ Cislighi, Gillespie, and Mackie, *Values Deliberation and Collective Action in Rural Senegal*.

⁴⁰ However, some norms may characterize networks that link different villages, so those subject to a norm could be greater than 500.

⁴¹ In the language of the public reason theorist, each is treated as a free and equal moral person. See Gaus, *The Order of Public Reason*, chap. 1.

dependent on ignorance or coercion, but on the reflective normative convictions of those they govern.

9.4. Accountability

We tend to interpret moral rules in ways that favor our own interests and, of course, we are often tempted to cheat on them. This can undermine a common morality, as each veers toward acting on self-interest. Thus a critical part of a functioning social morality is a practice of accountability, in which apparent violators are required to justify their actions to others. Cephu initially tried to defend his actions; after seeing that these defenses were dismissed, he then admitted his guilt to regain good standing in the group. In an important recent experiment, Erte Xiao has shown that having to present justifications for one's actions tends to render one more sensitive to the expectations of others.⁴² Because this accountability relation tends to enhance sensitivity to the empirical and normative expectations of others, it (i) helps keep people's understandings of the rules coordinated (we are constantly finding out what they expect of us) and (ii) checks our inevitable tendency to slide toward self-interested action and interpretations.⁴³

Our evolved morality fulfills a necessary role in human social life—we should understand it, and its practice of accountability, in light of that role, as a way for rather independently motivated agents, with their own aims and interests, to nevertheless enmesh themselves in intense cooperative networks. We should also understand the place of justification and our evolved capacity for reason-giving in the practice of accountability in a similar, functionality-focused light, as do Sperber and Mercier:

By giving reasons to explain and justify yourself, you do several things. You influence the way people read your mind, judge your behavior, and speak of you. You commit yourself by implicitly acknowledging the normative force of the reasons you invoke: you encourage others to expect your future behavior to be guided by similar reasons (and to hold you accountable if it is not). You also indicate that you are likely to evaluate the behavior of others by reasons similar to those you invoke to justify yourself. Finally, you engage in a conversation where others may accept your justifications, question them, and invoke reasons of their own, a conversation that should help you coordinate with them and from which shared norms actually may progressively emerge. Reducing the mechanisms of social coordination to norm abiding, mindreading, or a combination of these two mechanisms misses how much of human interaction aims at justifying oneself, evaluating the reasons of others (either those they give or those we attribute to them), criticizing past or current interactions, and anticipating future ones.⁴⁴

In this matrix of functions, defending oneself and one's reputation with reasons in the face of a charge of noncompliance is critical. As we have seen, in small-scale cooperative societies, one's reputation as a good cooperator is absolutely critical; Boehm and Alexander insist that selection by reputation was perhaps the chief force in the development of

⁴² Xiao, "Justification and Conformity."

⁴³ See also Tomasello, *A Natural History of Morality*, pp. 2, 39ff.

⁴⁴ See Mercier and Sperber, *The Enigma of Reason*, p. 168.

conditionally reciprocal cooperators.⁴⁵ Sperber and Mercier add that in such social contexts there were strong selection pressures for argumentative justification. As Sperber, Mercier, and Haidt all argue, the model of reasoning selected in these contexts is that of an advocate for one's innocence, not an impartial Kantian inner tribunal.⁴⁶ "The main role of reasons," according to Mercier and Sperber, "is not to motivate or guide us in reaching conclusions but to explain and justify after the fact the conclusions we have reached."⁴⁷ Of course, as we saw with Cephu, like any worthwhile advocate, in giving justifications one may see that the jury isn't buying it, and one may have to revise one's plea—when one justifies, one becomes sensitive to one's audience and their expectations. But, again as with Cephu, the first attempt is almost always to argue one's case, and to see how many one can get to accept one's plea. Most often, one's plea is not simply a disinterested opinion about what the rule requires—it just so happens that this opinion best supports one's wider goals. Sperber and Mercier call this a "myside" bias.⁴⁸

It is difficult to overstate the importance to an individual of effectively presenting his case. As Peter DeScioli observes, "In the arena of human conflict, moral judgment is a powerful weapon. Accusations of wrongdoing such as lying, infidelity, or blasphemy can mobilize a mob of aggressors against an opponent. Moral condemnation turns cold shoulders toward colleagues, sends terrorists after political cartoonists, and pulls nations into costly wars."⁴⁹ As we saw with Cephu, in small-scale societies the costs can be high—starting with ridicule, then ostracism and perhaps expulsion.⁵⁰ Developing effective argumentative skills was thus a critical adaptive challenge to those living in human moralistic communities. "In this social environment, it is easy to see why natural selection would favor psychological adaptations for using moral judgment against opponents, and for avoiding prohibited actions to escape the wrath of condemners."⁵¹

§10 The Other Side of Morality

10.1. The Side-Taking Hypothesis

As do most ethnographers and evolutionary theorists, I have been stressing the coordinating, cooperative role of morality in the evolution of human ultra-sociality. I do

⁴⁵ Boehm, *Moral Origins*, Chap. 7; Alexander, *Human Social Evolution*, Chap. 9.

⁴⁶ See Mercier and Sperber, *The Enigma of Reason*, p. 124; Haidt, *The Righteous Mind*, pp. 81ff. On the Kantian view, see Hill, "Four Conceptions of Conscience."

⁴⁷ Mercier and Sperber, *The Enigma of Reason*, p. 112. We need not follow this line of thought to a comprehensive debunking of impartial reasoned argument; in contexts of accusation and justification, their analysis of argumentative reasoning is powerful.

⁴⁸ *Ibid.*, pp. 218ff.

⁴⁹ DeScioli, "The Side-Taking Hypothesis for Moral Judgment," p. 24. Cf. Kurt Baier, "Moral talk is often rather repugnant. Leveling moral accusations, expressing moral indignation, passing moral judgment, allotting the blame, administering moral reproof, justifying oneself, and, above all, moralizing—who can enjoy such talk? And who can like or trust those addicted to it?" *The Moral Point of View*, p. 1.

⁵⁰ The proclivity to punish others for norm violation arises very early—it is present at three years of age. Tomasello, *A Natural History of Human Morality*, p. 101.

⁵¹ DeScioli, "The Side-Taking Hypothesis for Moral Judgment," p. 24.

not think it would be an exaggeration to claim that the development of a shared, social morality is probably the most important milestone in our journey from dominating Machiavellian strategic primates to ultra-social cooperators.¹ We are close to having assembled what might be called the Modern Egalitarian Moral Package, which includes sustained subordinate rebellion—protection of autonomy via reverse domination of would-be dominators—conditional cooperative motivations, a concern with fairness, ability to learn and internalize cooperative rules that treat members as free and equal, sensitivity to social rules, a flexible conscience, reconciliation on common rules, readiness to punish offenders, and a practice of accountability that requires argumentative skills. This last, we have seen, also lays the foundation for divisive arguments: if Cephu could have gotten away with it, he would have rallied others to his side. Cephu eventually conceded his guilt only because no one took his side—even those in his small band were not willing to rally to his defense. As the case of Cephu so clearly illustrates, practices of accountability seek to rally side-taking. When charged with a violation, we make our appeal to third parties.

In a recent series of papers, DeScioli and Robert Kurzban have stressed this “side-taking” view of morality.² We have seen that humans seem preadapted to coalition formation (§6.3). When disputes arise, third parties take sides: coalitions are encouraged to form around the different sides as each disputes responsibility.

When a bystander chooses sides in a conflict, an important consideration is which side other bystanders will support. We refer to the bystanders or outsiders to a conflict as “third parties,” distinguishing them from the two initial parties between whom the dispute began. We assume that third parties incur greater costs from being on the losing side than the winning side and, further, that numerical superiority provides an advantage. These two factors together give rise to an important adaptive problem: avoiding being on the minority side. Third parties need to anticipate which side the majority will take to avoid being outnumbered and suffering a costly defeat. When all third parties seek to side with a majority, they collectively face a coordination problem that requires synchronizing their side-taking decisions.³

On the side-taking hypothesis, morality is a way of taking sides that helps distance the third parties from any ties (such as kinship) that they may have to one of the disputants. When individuals get into disputes, if each side’s kin rallies to their own, the group will tend to break down into conflict. We thus can see why impartiality is so central to morality: over the long run, impartial rules do not favor a particular kin group, and thus allow third parties to coordinate their judgments and avoid

¹ Another critical evolutionary development that led humans away from the other existing great ape lines was terrestriality: *Australopithecus*, around 4 million years ago, had abandoned tree living; earlier *Ardipithecus ramidus* (6–7 million years ago) lived at least partially on the ground. As the *Homo* genus adapted to terrestrial life and moved into the savannah, a variety of evolutionary changes occurred. Of first importance was that life became focused on group hunting. Suwa et al., “The *Ardipithecus ramidus* Skull and Its Implications for Hominid Origins.”

² DeScioli and Kurzban, “A Solution to the Mysteries of Morality”; DeScioli and Kurzban, “Morality Is for Choosing Sides”; DeScioli, “The Side-Taking Hypothesis for Moral Judgment.”

³ DeScioli and Kurzban, “A Solution to the Mysteries of Morality,” pp. 480–1.

intra-group conflict.⁴ Again, the tale of Cephu is instructive. He clearly broke the rule against cheating in net placement: his band was not prepared to come to his aid, and his wife acquiesced to having her meat expropriated.

10.2. Impartiality, Objectivity, and Side-Taking

DeScioli and Kurzban's analysis also applies to disputes *about* morality. Suppose that a group is debating about two alternative moral rules. A critical factor in determining which rule will be more effective in coordinating cooperative behavior is deciding which alternative is most able to rally third parties and so avoid costly intra-group conflict. As John Tooby and Leda Cosmides argue, this crucially will be about which rule is more impartial—which rule can be endorsed by a more extensive majority. If some rule is to, as Tooby and Cosmides put it, “climb the ladder of increasingly wide support,” it must be endorsed by as many of the group members as possible, and so must be more impartial than the alternative.⁵

The underlying idea is that a community has available a number of possible moral equilibria—rules which all can endorse, given that others endorse them.⁶ On this view, communities often face a coordination problem along the lines of Display I.3.

		Betty	
		R_1	R_2
Alf	R_1	1 st 2 nd	3 rd 4 th
	R_2	4 th 3 rd	1 st 2 nd

Display I.3. The coordination-reconciliation game.

Here R_1 and R_2 are alternative moral rules. For the many reasons canvassed thus far, players Alf and Betty have strong reasons to coordinate. But, as we have seen, they have different interests and thus, to secure coordination, they must reconcile. Now one way to reconcile is if, because it is more impartial, R_1 has many more adherents; Betty, seeing that it is impossible to draw others to her favored R_2 , may decide to reconcile on R_1 . P. Kyle Stanford argues that another way to secure coordination is a belief that R_1 is objectively correct: “We do not merely affiliate with those who share our own moral commitments against those who do not—in addition, we think that we are right to do so and would be wrong to do otherwise.”⁷ There are certainly cases in

⁴ “Over the long run” is important: our concern is not individual cases, but the long-term dynamics of impartiality over many cases. In the short run, almost any outcome is possible.

⁵ Tooby and Cosmides, “Groups in Mind: The Coalitional Roots of War and Morality,” p. 220.

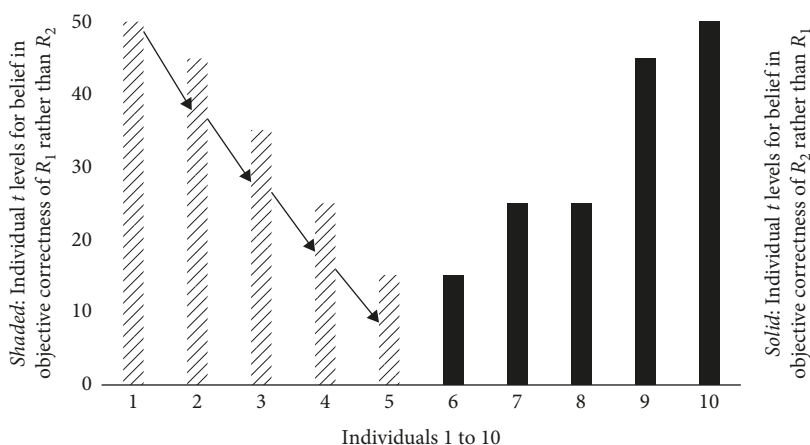
⁶ On rules and social equilibria, see Gaus, *The Order of Public Reason*, chap. 7.

⁷ Stanford, “The Difference between Ice Cream and Nazis,” p. 5.

which a conviction about the objectivity of morality can lead to convergence on an equilibrium. Consider a simple model, supposing:⁸

- (i) Alf endorses the objectivity of judgment and presently upholds R_1 .
- (ii) Alf has some threshold t which determines the percentage of the population that must also view R_1 as objectively correct if he is to continue doing so. Each person, we assume, has some such *other-regarding evidentiary threshold*: he can only continue to uphold the objective correctness of his belief in R_1 if his personal t threshold is met. I assume that other considerations enter into his conclusion: the number of others also endorsing R_1 is not his only evidence for its correctness.
- (iii) Some individuals may persist in believing R_1 (or, alternatively R_2) even if only a small percentage of others also do: their t values may be quite low. All that is required is that, if one believes in the correctness of R_1 (or, alternatively, R_2) there must be some threshold t of others that must also believe R_1 . A majority of others agreeing is not necessary. This allows us to model rational difference within a group, where each half of the group believes in the objective correctness of its favored rule.
- (iv) If one's threshold is not met, one will not believe R_1 . To make things simple, one then switches to R_2 .

Display I.4 helps clarify the dynamics of an objectivity-induced convergence.



Display I.4. An objectivity-induced cascade.

Patterned lines in Display I.4 indicate R_1 believers, solid lines R_2 believers; at the outset, society is equally split between those who uphold the objective correctness of R_1 and R_2 .

⁸ I am grateful for discussions with Shaun Nichols on these matters. See also Ayars and Nichols, "Rational Learners and Metaethics."

Suppose that person 1 either changes her mind, and decides that R_2 is correct, or determines that she has overestimated the number of people who endorse R_1 . In any event, she no longer endorses R_1 , and becomes an R_2 advocate. But now we see that person 2's threshold is not met: he requires that 45% of the group endorse R_1 (it is now at 40%), so person 2 thus switches, now leaving person 3 below his threshold of 35%, and so on to person 4, and then to 5.⁹ The core idea here is that belief in the objectivity of morality makes one sensitive to what others endorse, and so belief cascades of this sort can produce a convergence on a common rule.

However, another feature of objectivity threatens this happy analysis: a belief about moral objectivity of R_1 is associated with withdrawing cooperation from those who reject (what we believe to be) objective morality.¹⁰ These others, after all, are not just doing something different—they are acting IMMORALLY. If society is divided between R_1 and R_2 proponents, should each refuse to cooperate with those others, group cooperation is endangered. So beliefs in objectivity in deeply divided societies would *prevent* convergence on an equilibrium: if people are ready to infer that a rule is objectively correct in the face of considerable disagreement, conflict would be exacerbated, and cooperation stymied.

Interestingly, recent work indicates that when a community is significantly divided on some issue it tends *not* to be viewed as one that has an objectively correct answer, but as the overwhelming majority comes to adopt a position, belief in its objectivity increases.¹¹ If the belief in the objectivity of morality is characterized by very high threshold (t) values in this way, it would seem more of a device for *stabilizing* moral equilibrium than generating one (as in the preceding model). In such “high t value” cases, should an R_2 equilibrium be established, a belief in the moral objectivity of R_2 , and thus condemnation of other alternatives, would help secure the R_2 equilibrium, since people will come to believe it is not only what we have agreed upon, but it is the objectively correct answer. “Each society believes that its behavior is appropriate, while its neighbors do things improperly.”¹² This supports the hypothesis that belief in moral objectivity evolved as an *equilibrium stabilizing mechanism*.

10.3. Putting Some Dynamics Together

A Strategic Model of Justification

The Basic Model

We have arrived at a rather complex view of the place of justification in our evolved human morality. Cooperation requires a shared morality; such morality must ground shared expectations that others will behave in a predictably cooperative way; because our interests interact with our judgments, we tend to use any ambiguity in the rules

⁹ Other interpersonal evidence-based dynamics can lead to agreement. See Muldoon et al., “Disagreement behind the Veil of Ignorance.”

¹⁰ Stanford, “The Difference between Ice Cream and Nazis,” pp. 7–8. See also Skitka, Bauman, and Sargis, “Moral Conviction.”

¹¹ Ayars and Nichols, “Rational Learners and Metaethics.”

¹² Flannery and Marcus, *The Creation of Inequality*, p. 55.

to favor our own interests; and, of course, cheating is an ever-present possibility. We thus need to reconcile and accept common rules that are justified to all participants in the sense that they adequately align with people's normative convictions and interests. Yet self-interest and disagreement always threaten instability in the shared rules. As a result, it is necessary for us to police each other's performances and hold each other accountable for unjustified violations; and so a critical human cognitive capacity is to justify oneself to others and show that one should not be held accountable (and possibly punished) for wrongful violations. It is worthwhile stressing how dangerous punishment can be to a person's well-being. In all societies, those deemed to have violated the group's moral rules are, at best, downgraded as possible cooperative partners, and often are subject to various degrees of ostracism. It would be quite remarkable if, given all this, in contexts of accountability justification did not take on the role of advocacy for one's position. But this in turn leads to morality as a side-taking phenomenon. When charged with violations, people make out cases that they are not to be held accountable—sometimes by conceding and pleading extenuation, sometimes by challenging the accepted rule.

When, we might wonder, will taking sides—and attendant claims to impartiality and objectivity—lead to stabilization on a common rule, and when will it produce protracted disagreement? At this point it will be useful to begin to put together some of the various dynamics of human morality that we have been exploring via a simple model (that we will further develop in Part II). This very simple model has three key assumptions.

- (I) *Moral Disagreement*: We suppose that people often differ on what would be the morally best cooperative rule. Because we suppose that moral views and self-interests are intertwined, we suppose that a person would gain more overall utility from universal action on her favored rule than the alternative. A person's decision as to what rule is best, we assume, depends on her view of what is morally correct; but since moral judgment is informed with self-interest, and people's justifications track their interests, we can expect people to justify different rules.

We can clarify this a bit through some formalization. Let $\mu_A(R_i)$ represent Alf's judgment of what we might call the "inherent" utility of rule (i); suppose this is a cardinal, interpersonally noncomparable measure. To describe this as utility is simply to say that Alf has a scale of choice worthiness for moral rules.¹³ We have seen that disagreement about the choice worthiness of various moral rules is inherent in the nature of human moral order (§9.3).

- (II) *The Benefits of Expanding Cooperation*: The morality-as-beneficial-cooperative framework view is supposed: every individual sees increasing her network of moralized cooperation as always, to at least some extent, a good thing—and conversely, being excluded from a network is always, to some extent, a bad thing. Expanding cooperation is by no means all that a moral agent is interested

¹³ This idea is further explained in section 20.

in, but it is always an interest. This is the classic evolutionary view, and we are trying to better understand where it might lead us. Individuals thus have two bases for deciding whether to act on R_1 : its “inherent utility” [$\mu_A(R_1)$] and how many others are acting on it. Call $U(R_1)$ this total utility.

It is important that this by no means implies that individuals’ $U(R_1)$ increases in a linear way with the number of others acting on R_1 . Typically, we would suppose that most would place little additional value on the move from a tiny to a slightly larger cooperative network: we would expect individuals to have some effective threshold, such that until the network is of some minimum size, that person would place little value on it.¹⁴ However, we can be ecumenical about this, allowing a variety of functions relating increased numbers in a network and increased valuing of participation.

(III) *Strategic Advocacy*: This leads us to our third assumption, DeScioli and Kurzban’s strategic view of moral disagreement: “Individuals stand to gain by proposing and defending moral rules that benefit themselves”¹⁵ and thus will argue strategically in moral contexts. Consistent with this and Mercier and Sperber’s view (§9.4), it is supposed that a person’s response to a charge of non-compliance is to be explained by what, all things considered, best advances her utility (with this including moving society to her most favored moral position).

Consider, then, a simple world composed of group G with two possible moral rules, and everyone acts on one or the other. Suppose that Alf has performed action ϕ . Betty, a follower of moral rule R_1 , claims that ϕ is wrong on her interpretation of R_1 , and she holds him responsible for its violation. In our simple model Alf has only two options. He can CONCEDE guilt, accepting that Betty’s is the proper interpretation of R_1 , a correct moral rule. Alternatively, he can DEFEND, rejecting the rebuke, advancing a justification claiming either (i) that Betty’s is not the proper interpretation of R_1 or (ii) that R_1 is not the correct moral rule. Let us combine i and ii by saying that, when he DEFENDS, Alf claims that ϕ was not wrong according to moral rule R_2 , which is either the correct interpretation of R_1 or the correct moral alternative to R_1 .

If Alf DEFENDS, he will be excluded from the R_1 cooperative network. He will have sent a signal to the upholders of R_1 that he dissents, and so that he cannot be relied upon to act on the expectations generated by R_1 , and that he refuses to be held accountable for its violation. As Tomasello indicates, this may be a core function of moral justification: “moral justification is not necessarily aimed at producing moral actions or achieving empirical accuracy, it is aimed at finding shared values that demonstrate one’s continued identification with the moral community.”¹⁶ Given Assumption II, the larger the R_1 network (the “moral community”), the higher the costs of nonparticipation. On the other hand, by defending R_2 , Alf signals his loyalty to the R_2 network (all those in the group G who are not in the R_1 network). Of course, size of network is not decisive: if in Alf’s judgment the inherent utility of R_2 [$\mu_A(R_2)$]

¹⁴ See Bicchieri, *The Grammar of Society*, pp. 11, 223.

¹⁵ DeScioli and Kurzban, “A Solution to the Mysteries of Morality,” p. 488.

¹⁶ Tomasello, *A Natural History of Morality*, p. 114.

is high, a cost of CONCEDEING would be remaining in the R_1 network, which offers less favorable terms of moral cooperation [i.e., $[\mu_A(R_2)] > [\mu_A(R_1)]$, though if R_1 network is sufficiently larger, that can compensate). If he DEFENDS and so joins the R_2 network, the moral superiority of R_2 would be an additional benefit (which could compensate for it being a smaller network). His decision will thus be driven by the overall relative benefits of R_1 and R_2 to him, and the relative size of the cooperative moral networks that provide these benefits. Alf, then, will choose:

DEFEND if: $U_A(R_2) > U_A(R_1)$;

CONCEDE if: $U_A(R_1) > U_A(R_2)$.

Case 1: Overwhelming Choice as Stabilizing an Equilibrium

Consider three paradigmatic cases. In the first, Alf regularly concludes that $U_A(R_1)$ is much larger than $U_A(R_2)$, as does just about everyone else. Moral diversity (Assumption I) still holds (it still may be the case that $\mu_A(R_2) > \mu_A(R_1)$), but by far most accept R_1 , and so the size of the R_1 network is so overwhelmingly large that Alf does not want to be excluded. Here the members of group G overwhelmingly support R_1 , and so treat any violation as impermissible and warranting moral condemnation. In this case, even if Alf finds R_1 objectionable as a basis for social relations, the sheer number who support it will most likely lead him to CONCEDE, and not seek to justify his ϕ -ing. Here, as Stanford suggests, we should expect the group to believe in the objectivity of the R_1 rule. When a large majority of our group believes that R_1 is universally correct, it will not tolerate departures from it: to reject R_1 is to endanger one's status as a competent group member. In Case 1, accountability and justification function as aids to moral coordination. To DEFEND would constitute a public signal that one rejects what the group deems to be the demands of morality itself. There are few benefits to arguing for a revised rule in this accountability context, and great costs to explicitly breaking with what is understood as morally right. Insofar as Alf uses justification-as-advocacy, it will be in the way of CONCEDEING, say, to point to excusing factors.

Case 2: Choosing Sides and Convergence

In the second case, the group is divided into, say, three rough types: a subgroup of R_1 advocates; another who advocate R_2 ; and a significant "quasi-indifferent" Q group for whom $U_Q(R_1)$ and $U_Q(R_2)$ are either equal or very close to equal. Here, if Alf decides to DEFEND the morality of action ϕ , members of group G choose sides. If Alf DEFENDS, he will have two aims: (i) to signal to R_2 devotees that he is a reliable member of their network and (ii) to convince members of the Q subgroup that impartial considerations are on the side of R_2 , thus expanding the size of the R_2 network (and so his own $U_A(R_2)$). Many in the quasi-indifferent subgroup are apt to be less sensitive (§9.3) to either R_1 or R_2 ; in their eyes, neither manifestly better aligns with their normative convictions or, perhaps, their personal normative convictions render them ambivalent about the rules. In this case we would not expect the overall G group to have strong convictions about the moral objectivity of either rule. Here, should Alf defend, his use of justification-as-advocacy is likely to be *reformist*, arguing to others, especially to the wavering R_1 followers in the quasi-indifferent subgroup, of the superiority of R_2 and

the flaws of R_1 . Convincing them to also adopt R_2 would, *pro tanto*, increase $U_A(R_2)$ by increasing his weight to the R_2 rule. The existence of the less sensitive quasi-indifferent subgroup is critical for this case: they are most apt to be moved by arguments for or against R_1 and R_2 . In crafting his justification, Alf will be sensitive to their concerns. To employ Tooby and Cosmides's phrase, if R_2 is to "climb the ladder of increasingly wide support," the case for it must extend beyond its parochial appeal to a highly committed R_2 subgroup. This incentivizes presenting impartial and general cases for R_2 , appealing to a wider array of views. Thus we see that the critical features of moral argument—impartiality and universalization—are elements of strategic justification (Assumption III). In Case 2, moral reconciliation on a publicly justified shared moral rule is entirely possible, despite differences in moral views. Indeed—and this is the interesting point—such a society tends to produce justifications with wider appeal. We thus arrive at an important hypothesis that we shall expand upon in Part II: *societies that have successfully coped with moral diversity at one level may well be those that can continue expanding their moral networks because they have achieved wider-based, more impartial, justifications*. In "climbing the ladder" of wider appeal in a diverse society, they have crafted their rules to accommodate greater diversity.

Note here that the very justificatory competency that is critical to a stable shared moral rule (Case 1) also can be employed to undermine the current rule and move to a new publicly justified rule. Justification must be able to perform this destabilizing role if a cooperative moral system is to learn and adapt. As recent analyses such as Haidt's, Stanford's, and DeScioli and Kurzban's have recognized, any adequate account of morality must be able to induce change as well as provide stability.¹⁷ Moral diversity and conflict may be an engine of moral reform, pointing toward a new cooperative equilibrium.¹⁸ On the other hand, we should expect continued conflict on many matters. "As moral projects climb the ladder to broader audiences (being recast and potentially applied to increasingly broad sets of individuals), any given individual will be bombarded with increasing numbers of candidate moral rules."¹⁹

Case 3: Ossified Conflicting Choices

The dynamics considered in Case 2 explored the implications of *moderate* polarization. Suppose now that G is *radically polarized* into two groups: for one group $U(R_1)$ is much larger than $U(R_2)$, while for the other $U(R_2)$ is much larger than $U(R_1)$. There is no quasi-indifferent group. It is important to stress that the two subgroups need not be the same size; differences in the inherent valuation of the rules [$\mu(R_1)$ and $\mu(R_2)$] may compensate for lack of numbers in a person's decision-making.²⁰ Here the first group will always CONCEDE when faced with a violation of R_1 , the second will always DEFEND (and, *mutatis mutandis*, the opposite with R_2 violations). Here each subgroup strategically uses justification-as-advocacy to solidify the group's devotion to the rule, not to generate converts from the other: no one is trying to broaden their

¹⁷ See DeScioli, "The Side-Taking Hypothesis for Moral Judgment"; DeScioli and Kurzban, "A Solution to the Mysteries of Morality"; DeScioli and Kurzban, "Morality Is for Choosing Sides"; Stanford, "The Difference between Ice Cream and Nazis"; Haidt, *The Righteous Mind*, chap. 12.

¹⁸ See further my *Tyranny of the Ideal*, pp. 230–40.

¹⁹ Tooby and Cosmides, "Groups in Mind," p. 224.

²⁰ See Skitka, Bauman, and Sargis, "Moral Conviction."

justifications as a way of appealing to the (nonexistent) quasi-indifferent group (cf. Case 2). In DEFENDING or CONCEDED, one only signals devotion to one or the other faction. Each subgroup's justifications reinforce its own rules, while viewing the other as morally beyond the pale. "Myside" bias will be deeply rooted and reinforced. Each preaches to the choir, and each increasingly firmly holds a moral view that simply cannot be embraced by the other group.

If this division is overlapping such that a number of disputes all have the same subgroup boundaries while the other subgroup strongly opposes these rules, the empirical evidence highlights three possible upshots. The *first*, most dire (and not terribly unlikely), possibility is that inter-group hatred arises, not only hindering cooperation but inducing conflict and, at the extreme, violence.²¹ As I have stressed, when morality is evoked, the other group is not simply different: they are turning their backs on MORALITY. When one group has achieved consensus on a rule, it is apt to see this rule as objectively correct. Condemnations as not only "immoral" but also "evil" may arise. Moreover, as we have seen, any attempt to punish the other subgroup is likely to invoke counter-punishment. Here justification as choosing sides leads to sustained moral conflict, or even what Tooby and Cosmides describe as "moral warfare."²² Justification becomes purely a group loyalty-signaling device. The very mechanisms at the core of our evolved moral practices drive discord and impair cooperation

A *second* possibility is that the subgroups gravitate apart: as we have seen, given that a perception of another as immoral is associated with fewer cooperative relations with them: many such disputes are resolved by self-segregation.²³ As I have argued elsewhere, this may be a social rather than a spatial separation; the subgroups can carve out different social worlds in which, to some extent, their distinctive moralities hold sway. Interestingly, people's views of the morality of an out-group appear to become increasingly "relativistic" as cooperation with them decreases and they operate in different cultural milieus.²⁴

Lastly, the dispute may eventually be solved by *demoralization*. Evidence indicates that as people's empirical expectations about the behavior of others is repeatedly disappointed, the belief in the existence of *any* rule may be undermined.²⁵ To accommodate this in our model, we need to add a third option—no rule at all. In this case as, say, R_1 followers in G observe others (the R_2 followers) not acting on R_1 , their empirical expectations that R_1 is the "done thing" in group G may be undermined. Not able to easily distinguish the different groups by ethnic markers (§11.3), members of G may decide that no rule is sufficiently followed or can gain sufficient support, and so they opt for "no rule." This seems a familiar dynamic in the natural history of the moral order: when diverse populations cannot settle on a moral rule but cannot subsist apart, they may demoralize the relevant behavior.

²¹ Böhm, Thielmann, and Hilbig, "The Brighter the Light, the Deeper the Shadow: Morality Also Fuels Aggression, Conflict, and Violence."

²² Tooby and Cosmides, "Groups in Mind," pp. 224–5. Cf. Vallier, *Must Politics Be War?*

²³ Stanford, "The Difference between Ice Cream and Nazis," p. 9.

²⁴ See Sarkissian et al., "Folk Moral Relativism."

²⁵ Bicchieri and Xiao, "Do the Right Thing: But Only If Others Do So."

Convergence, Conflict, and Choosing Sides

The lesson of our model is that, given the dynamics of our evolved moral relations, human morality is characterized by moral conflict as well as consensus. Although the orthodox “morality as cooperation” account remains the core of the natural history of human morality, we see that traits that have evolved to stabilize morality as cooperation—such as argumentative justifications and choosing sides—can lead to conflict and self-segregation of groups, even when this impairs cooperation. A key factor underlying morality is not simply the imperative to coordinate, but that we disagree on how to do it, and so need to reconcile. Thus, since at least the Late Pleistocene era, humans have had to play the coordination-reconciliation game (Display I.3). Public justification, we have also seen, is critical in grounding shared norms to which the population is sensitive. Here, justification works to assist reconciliation. At the same time, we employ justification strategically to promote our own moral views and interests, and this can lead to conflict, as we rally others to take our moral side. Our strategic model indicates that whether such argumentative justification leads to convergence or sustained conflict critically depends on the population’s distribution of moral views, and the extent to which change occurs through impartial arguments that bring others to our side.

§11 Cultural Evolution

11.1. The Evolution of Moral Learning

We have been investigating dynamics that are deeply rooted in the human moral project. It is reasonable to suppose that these elements of the Modern Egalitarian Moral Package had evolved by 40,000 years ago. This evolved package, it will be recalled, includes sustained subordinate rebellion, protection of autonomy via coercive control of potential upstarts and deviants, conditional cooperative motivations, a concern with fairness, ability to learn and internalize cooperative rules that treat members as free and equal, sensitivity to social rules, a flexible conscience, reconciliation on common rules, readiness to punish offenders, and a practice of accountability that requires argumentative skills. It was traditionally thought that genetic evolution moves slowly, and so something near 40,000 (or perhaps, 20,000) years ago marks the effective end of a genetic evolutionary account.¹ Newer evidence indicates that significant genetic changes can occur much more swiftly than this. Indeed, it seems that human evolution has markedly speeded up in the last 50,000 years.² By far the dominant explanation of

¹ Until recently it was standardly assumed that something like 1,000 generations (or 25,000 years in humans) were necessary for major genetic changes.

² See Haidt, *The Righteous Mind*, p. 247. For a more radical view, see Cochran and Harpending, *The 10,000 Year Explosion*; Turchin, *Ultrasociety*.

recent human genetic change is gene-culture coevolution.³ Culture has largely determined the adaptiveness environment for human genetic success: given that rewards and punishments are explicitly targeted at certain sorts of culturally approved or disapproved behaviors, evolutionary pressures can be extremely strong.⁴ Reproductive opportunities tend to be denied to those who flout group norms. “Over our evolutionary history, norm violations such as ignoring a food taboo, botching a ritual, or failing to give one’s in-laws their due from one’s hunting successes meant reputational damage, gossip, and a consequent loss of marriage opportunities and allies.”⁵ Culture also is constantly increasing new evolutionary niches: as culture develops, new ways of being adaptive are developing along with it.⁶ Gene-culture evolution is thus “autocatalytic”: as culture develops, it drives genetic adaptations to it, which then provide the opportunities for new cultural developments, which create new pressures for genetic adaptation.⁷ Such a process has a self-reinforcing “runaway” dynamic, driving evolution quickly forward on a cultural path. It is thus profoundly misleading to contrast “nature” and “culture,” as if our nature has not been shaped by the fact that our evolutionary environment has largely been shaped by culture. Hayek was certainly right on this point: “Mind is as much a product of the social environment in which it has grown up and which it has not made as something that has in turn acted upon and altered these institutions.”⁸

Most of the moral developments we have been considering—especially those in the Late Pleistocene and early modern eras (say up to 10,000 years ago)—were the result of gene-culture evolution. Once humans developed rudimentary rule-guided cooperative practices (such as hunting), there were very strong culture-based pressures encouraging the success of those who were more adept at rule-based guidance and attendant moral justifications. We have, as Henrich puts it, domesticated ourselves into being better rule followers.⁹ Again Hayek was correct: “Man is as much a rule-following animal as a purpose-seeking one.”¹⁰ Though internalized—we see the rules as bound up with who we are—they did not emanate from within. The rules themselves are social adaptations—solutions to problems of group life, often problems that no one even realized were problems. “Social norms make it possible for humans to solve—often without anyone understanding how—what would otherwise be inescapable social dilemmas. Social life is riddled with opportunities to exploit others, which most people don’t even notice.”¹¹

³ Most importantly, as analyzed in the works of Boyd and Richerson. Their key theoretical work is *Culture and the Evolutionary Process*. See also Richerson and Boyd’s more popular presentation, *Not by Genes Alone*, as well as Laland, *Darwin’s Unfinished Symphony*, chap. 9; Mesoudi, *Cultural Evolution*, chap. 3; Henrich and Henrich, *Why Humans Cooperate*, chap. 2.

⁴ Laland, *Darwin’s Unfinished Symphony*, p. 216; Henrich, *The Secret of Our Success*, p. 52. This is critical to Boehm’s social selection hypothesis in *Moral Origins*.

⁵ Henrich, *The Secret of Our Success*, p. 5.

⁶ Laland, *Darwin’s Unfinished Symphony*, p. 125.

⁷ See Henrich, *The Secret of Our Success*, pp. 57–64; Laland, *Darwin’s Unfinished Symphony*, pp. 29, 148–50, 189–90, 230–1. See Part II, §15, in this volume.

⁸ Hayek, *Rules and Order*, p. 17.

⁹ Henrich, *The Secret of Our Success*, pp. 318ff.

¹⁰ Hayek, *Rules and Order*, p. 11.

¹¹ Henrich, *The Secret of Our Success*, p. 145.

Our natural history does not end at gene-culture coevolution. Laland holds that there have been three main epochs in human evolution: when genetic evolution predominated; when gene-culture evolution did; and now, when the main driver is cultural evolution.¹² And of course it was cultural evolution that Hayek stressed.¹³ Given this, it is surprising indeed that so little of the work on the evolution of morality focuses on cultural evolution.¹⁴ While our capacity to be guided by social norms arose through gene-culture evolution, the norms themselves are predominantly a product of cultural evolution. Moral rules are a cultural phenomenon.

As Henrich (in a distinctly Hayekian vein) stresses, we do not understand the function of numerous culturally evolved social rules—they are causally opaque to us. Consider one of Henrich's case studies: the custom of using of fresh wood ash in the preparation of corn dishes by the Mapuche.¹⁵ On the face it, this looks like a sort of food contamination. It was not adopted by North Americans and Europeans when they took up the use of cornmeal as a low-priced staple. After they did so, however, pellagra became common—a disease that, in the first part of the twentieth century, killed millions of the poor. Since the disease arose in the spring, it was assumed that some form of contamination during the winter had occurred, but measures to improve storage techniques has no effect. As far back as 1915, one doctor hit upon the correct causal account: a strict diet of cornmeal (such as over the winter when other foods were not available to the poor) leads to niacin deficiency, the cause of pellagra. Mixing ash, which has lye, an alkali, allows the niacin to be unlocked. But this explanation went against the dominant contamination model of food-borne diseases, and actually seemed to endorse contamination. Thus the best science of the time rejected the correct causal model for decades. The Mapuche did not have the correct causal theory, but they did have cultural knowledge that tracked it.

Perhaps while recipes are opaque, norms stating injunctions and prohibitions are not. Consider another example from the work of Joseph and Natalie Henrich.¹⁶ On the Yasawa Island in Fiji there are rules against pregnant and nursing women eating certain large seafood such as eel, barracuda, large grouper, and shark, which form an important part of the population's diet. Studies show that ciguatera fish poisoning—a toxin that accumulates up the food chain, achieving its highest levels in large predatory fish—is common on Yasawa Island. Ciguatera poisoning is often serious and occasionally fatal. Henrich and Henrich's study revealed that the strict norms against eating such large fish obtain during pregnancy (with somewhat looser norms obtaining during breastfeeding) tracked the level of potential threats

¹² Laland, *Darwin's Unfinished Symphony*, pp. 234–5.

¹³ Hayek, like Laland, insisted that in our present age cultural evolution dominated. Because social evolution “differs from genetic evolution by relying on the transmission of acquired properties,” he writes, “it is very fast, and once it dominates swamps genetic evolution.” Hayek, *The Political Order of a Free People*, p. 156.

¹⁴ In *The Ethical Project*, Kitcher has a short discussion of cultural evolution, and acknowledges that biological and cultural success need not be linked (p. 109). But overwhelmingly, his story is about the egalitarian nature of the ethical project, an egalitarianism that has its roots in the period from 200,000 to 40,000 years ago. In *The Evolution of Morality*, Joyce follows the same pattern; with an occasional nod to cultural evolution, the evolution of morality is essentially about natural selection.

¹⁵ Henrich, *The Secret of Our Success*, pp. 102ff.

¹⁶ Henrich and Henrich, “The Evolution of Cultural Adaptations.”

of ciguatera poisoning. Given the serious health risk posed by ciguatera poisoning to fetuses and newborns, Henrich and Henrich concluded that the prohibitions are manifestly adaptive. Strikingly, however, although women display consensus on what the rules are, when asked to explain the purposes the rules serve the responses widely varied: women had learned the rule but its concrete purposes were obscure.¹⁷ In many ways, the essence of cultural existence is learning and internalizing rules without being clear what their function is, or the causal models that would show they secure desired aims. As cultural evolution proceeds, we are apt to find that the purposes of any given moral rule in our social life is very often obscure.

Now if culture practices and norms generally embody more information than the causal models of individuals, gene-culture evolution should have strongly selected for the ability to learn from cultural cohorts. What we cannot invent for ourselves we can learn from others. Thus Hayek's claim: "the brain is an organ enabling us to absorb, but not to design culture."¹⁸ As we saw earlier (§6.4) experimental results provide evidence for this conjecture. Chimps' causal reasoning is about the same as human toddlers, but the toddlers are far superior in their ability to copy others. Because humans copy a demonstrator's instrumentally useless as well as effective actions, humans have been said to demonstrate "over-imitation."¹⁹ This may make us seem like rather dumb imitators, but experiments also indicate that as children observe different demonstrators, some of whom do not employ useless actions, over-imitation rapidly decreases.²⁰ The fundamental point is that, because we are cultural creatures, humans are, first and foremost, learners: to successfully acquire the tremendous store of information within our culture, we must decide whom we are to copy. We tend, for example, to copy high-prestige persons, those who are more successful, or the majority.²¹ These copying mechanisms are the way in which cultural information is transmitted from generation to generation, and thus are critical to the evolution of culture. Without a transmission mechanism, no matter how advantageous, a cultural practice or innovation could not evolve.

11.2. Copying and Creativity

One of the most vexing questions in cultural evolution is the relative roles of innovation (or creativity) and copying (imitation). Both are necessary: innovators explore new ways of doing things, which are then taken up by copiers. As a recent model proposes, "cultural transmission (e.g., copying) reduces the loss of culture, but only creativity can produce new culture."²² This is consistent with the well-known "exploration-exploitation" trade-off model: groups and organizations need some who

¹⁷ Henrich, *The Secret of Our Success*, p. 101.

¹⁸ Hayek, *The Political Order of a Free People*, p. 157.

¹⁹ Henrich, *The Secret of Our Success*, pp. 108ff. Tomasello hypothesizes that such imitation is primarily a group identification device. *A Natural History of Morality*, p. 89. See §11.3 in this volume.

²⁰ Laland, *Darwin's Unfinished Symphony*, pp. 52–3.

²¹ Henrich and Henrich, *Why Humans Cooperate*, chap. 2; Richerson and Boyd, *Not by Genes Alone*, pp. 68ff; Laland, *Darwin's Unfinished Symphony*, p. 57.

²² Enquist et al., "Why Does Human Culture Increase Exponentially?" p. 52.

look for new ways of doing things (“exploration”), while others employ these insights and so gain the social benefits from them (“exploitation”). If too much energy goes into exploration, we are finding better and better ways of doing things, but not sufficiently reaping their benefits; if too many people exploit rather than explore, we are stuck where we are and cannot improve our practices and institutions. Copiers, after all, seem basically parasitical on the innovators; alone they cannot learn anything new. It seems intuitive that a group would settle on an equilibrium between innovators (who supply new information) and copiers (who exploit it).²³

As usual, things are more complicated. In evolutionary simulations and models it seems quite difficult to show that an agent benefits from being an innovator. An especially important simulation tournament—inspired by Robert Axelrod’s famous Prisoners’ Dilemma tournament (§8.1)—was conducted by Laland and his colleagues.²⁴ In this tournament, strategies confronted a “100-arm bandit”: a procedure with 100 possible behaviors and payoffs. Fitness was measured in the gains from the choices about which “arms” to pull. Agents (programs) confronted three choices: EXPLOIT, OBSERVE, and INNOVATE. EXPLOITING involved using the agent’s current knowledge to decide which “arm” it was best to pull. OBSERVING involved forgoing exploiting on a move to copy the actions of some other players, while INNOVATING involved forgoing EXPLOITING on a move and instead obtaining new, accurate information about one of the options (randomly determined). The behavior of the arms changed with a certain low probability to model the possibility that information might become outdated. More than 100 teams submitted strategies, seeking to optimize the mixture of EXPLOIT, OBSERVE, and INNOVATE choices. The winning entry—and indeed the better strategies generally—(1) mostly exploited, and only learned at intervals, and (2) when learning was employed it was only OBSERVE, not INNOVATE. The winning entry discounted past information and sought to anticipate when new knowledge would be useful. What is fascinating is that when this winning entry was modified such that when it did learn it INNOVATED rather than OBSERVED, its rank went from first to last, despite the sophisticated way in which it decided when to learn. The results of Laland et al.’s simulation demonstrated the tremendous advantage to the agent in combining lots of action with social learning based on copying.

It is important that the best strategies did not copy indiscriminately and stupidly: when to copy, and when to change copying behavior, was critical. Laland notes that the best agents “in our tournament possessed a repertoire of behaviors that they exploited flexibly. Following environmental change, successful strategies . . . would not just stick with outdated behavior, but switch to the behavior in their repertoire with the next highest payoff. In turn, when agents played EXPLOIT, other copying agents playing OBSERVE also acquired a behavior with a reasonable return. Consequently, unlike in many models, the copiers were not entirely dependent on the innovators to change behavior when the environment changed.”²⁵

²³ Rogers, “Does Biology Constrain Culture?”; Boyd and Richerson, “Why Does Human Culture Increase Human Adaptability?”

²⁴ Rendell et al., “Why Copy Others? Insights from the Social Learning Strategies Tournament.” See also Laland and Rendell, “Cultural Memory.”

²⁵ Laland, *Darwin’s Unfinished Symphony*, p. 72.

One of the counterintuitive features revealed by the study is that a population of copiers will tend to be more diverse than a population of innovators. The extent of diversity—ways of doing things—is a function of the entry of new behaviors *and* the loss of current behaviors. Diversity is decreased when existing patterns of behavior are lost: extensive copying of others helps prevent such cultural loss. In some well-known cases, the toolkits available in a population were lost by an accident (the death of older members), thus radically reducing the diversity and well-being of the culture.²⁶ High fidelity copying ensures that the current repertoire of behaviors is maintained, thus protecting diversity.²⁷ Moreover, when a society has a diverse repertoire of behaviors, when the environment changes rendering some behaviors less effective, there will be others ready to try. To be sure, sustained diversity requires new behaviors to be introduced. Without the introduction of new behaviors, social learning would eventually induce conformity as more and more copiers end up playing similar optimizing strategies. However, this need not be supplied by explicit innovation—imperfect copying (mistakes in learning) can also perform this role. Thus a fascinating hypothesis: while *high*-fidelity copying is necessary to reproduce existing cultural artifacts, if the copying is of *high*, but not *perfect*, fidelity it can be a source of new ideas. None of this shows that explicit innovation has no role, but it may well be more modest than we are apt to think.

11.3. Coordination, Social Learning, and Ethnic Markers

Recall that in the coordination-reconciliation game from Display I.3 there is more than one endorsable rule, and the task for a group is to reconcile on one. Our strategic model (§10.3) suggests that the dynamics of justification may sometimes lead to a stable coordination while, at other times, groups may be unable to reconcile on a common rule, dividing into subgroups. For present purposes—and remember that our focus is now on cultural evolution—what is important is that a member of a group learn the rule characterizing *her* group. If a social milieu mixes people from various groups (or subgroups), it is critical for successful coordination that each be able to clearly identify which moral rules and social norms are employed by one's group—not all groups employ the same norms in their interactions. That human societies are faced with multiple moral equilibria is one basis of “cultural relativism.”²⁸ Even if all humans saw roughly the same set of rules as morally eligible, they could end up coordinating on markedly different ethical codes, given their different solutions to the coordination-reconciliation game. Thus, although all humans possess the same Modern Egalitarian Moral Package, as they culturally evolve they will come to disagree on morality.

Because different groups reach different equilibria, successful moral coordination must depend on learners imitating those within their group rather than outgroup members. This often leads to the use of obvious “markers” such as tribal markings or

²⁶ Henrich, *The Secret of Our Success*, pp. 218ff., chap. 12.

²⁷ Laland, *Darwin's Unfinished Symphony*, pp. 73–4.

²⁸ For a still excellent analysis, see Westermarck, *Ethical Relativity*.

ethnicity, identifying who is in one's norm network. Now we should not jump from this to the supposition of a stark "us" vs. "them" moral distinction. Forager societies are not typically, as they are so often depicted, "closed" with little contact with outsiders. Ethnic markers and mixing with strangers go hand in hand. Foragers often develop extensive networks of dealing with neighboring groups, including ways of assisting them: mixing with non-band members is part and parcel of many foragers' existence.²⁹ Boyd and Richerson model the ease with which ethnic markers arise when learning is imitative and coordination on norms in a non-uniform population is critical: in such cases, "systems of moral norms ... should come to be marked by ethnic markers."³⁰ This will be especially so when different groups come together or interact: increased encounters with other groups should induce *greater* sensitivity to ethnic markers. Given the importance of successfully coordinating with others, we should expect humans to be sensitive to easily identifiable markers that distinguish models-to-be-imitated from to-be-ignored background behavior. Thus learning shared moral rules (the core of cultural evolution) and ethnic markers/stereotypes would seem to go hand in hand.³¹ Bicchieri et al. have recently shown that social proximity—whether, for example, certain obvious traits are shared—can be critical in preventing the erosion of norms in the light of knowledge of others' noncompliance.³² Not surprisingly, the existence of such ethnic markers and their effect on pro-social behavior are among the best-documented phenomena in anthropological research.³³ Psychological evidence indicates innate tendencies in this direction: very young children preferentially learn from those who share their mother's dialect.³⁴

This is all a consequence of our evolved nature as cultural beings. However, none of this implies that humans are inherently small-group minded or xenophobic—much less racist.³⁵ That we are alert to cues that reveal whether a person shares our moral network—and so is an eligible subject of emulation and trust—does not mean that we latch on to the same markers in all contexts. As social orders become diverse and multifaceted, individuals participate in various norm networks with diverse memberships. One might be a traditional Roman Catholic, a vegetarian, a doctor, a fanatical Buffalo Bills fan, and a third-generation Pole. Each of these can involve distinct norm networks, with different markers of membership in good standing: interestingly, in such a diverse society one may become *more* sensitive to markers, since different networks have different markers. When these networks are overlapping—if most Poles

²⁹ Flannery and Marcus, *The Creation of Inequality*, pp. 20, 27, 29, 34, 59, 548.

³⁰ Boyd and Richerson, "Shared Norms and the Evolution of Ethnic Markers," p. 129.

³¹ Henrich, *The Secret of Our Success*, pp. 44ff, 200.

³² Bicchieri et al., "Social Proximity and the Evolution of Norm Compliance."

³³ As Bicchieri et al. point out (*ibid.*), studies have shown that "individuals are more altruistic, trusting, and trustworthy towards in-groups rather than out-groups; more likely to comply with requests from by [*sic*] an in-group rather than an out-group; and more likely to take an in-group's advice into account when judging one's own actions compared to an out-group's." See also Tomasello, *A Natural History of Morality*, p. 136.

³⁴ Henrich, *The Secret of Our Success*, p. 44. However, Over and Carpenter ("The Social Side of Imitation," p. 8) argue that cultural learning is not itself an adequate explanation: in addition, selective imitation, they argue, signals group membership.

³⁵ Though racism may well be built on the psychology for employing ethnic markers. Henrich, *The Secret of Our Success*, pp. 204ff.

were also traditional Roman Catholics, vegetarians, doctors, and fanatical Buffalo Bills fans, then being entirely cued to Polish ethnic markers would make sense—and of course in more homogenous societies this may be approximately true. A worry about our contemporary societies is that in some ways we are reinventing the overlapping, simple networks of earlier times. The markers for Democrats (owns a Prius, couples have different last names, apologizes for eating meat) and Republicans (owns a pickup, goes to church, has a gun) identify memberships in a number of reinforcing norm networks and social groups. However, note also that in modern societies many of these “ethnic markers” are voluntary. For much of modern European history, dress was a marker of one’s moral—often class—network. As late as the mid-twentieth century, clothing clearly indicated middle-class, working-class, or bohemian identities, values, and social groupings.

Again, the proper conclusion is that we are complex, often somewhat ambivalent, cooperators and moralists. On the one hand, we generally gain by expanding our moral networks of cooperation to include more potential partners. The moral imperative to devise impartial justifications and principles can be understood, from this evolutionary perspective, as a device to expand our cooperative network by getting more on our side (§10.2). On the other hand, the very essence of solving the coordination-reconciliation game and cultural evolution is to identify those who coordinate *our way* so that we can learn *our* rules from them. This is reinforced by our need to trust others to conform to moral rules—who will confirm our empirical and normative expectations about their behavior. It is critical to know who is playing the same coordination option, and for that we search for obvious markers. But that we are cued to markers to help us distinguish who is in our moral network from who is not does not mean that we are cued to a uniform distinction between “us” and “them”—and it certainly does not mean that we are “tribal” (§§12.1, 13.4).

11.4. The Mechanisms of Cultural Evolution

As I have said (§5), for the most part I wish to put aside questions about the evolutionary mechanisms for the developments I have been canvassing. Again, this is not because these mechanisms are unimportant to an evolutionary analysis (anything but!); however, the disputes between them often turn on complicated modeling issues—and it seems that in the end the proper analysis will invoke several different mechanisms, sometimes at different points in the evolutionary history and sometimes operating simultaneously. However, as we saw in section 1, since Hayek—and indeed most theorists of cultural evolution—stressed “group selection models,” it is important to try to get clear about the relation between these models regarding genetic evolution, genetic-cultural coevolution, and cultural evolution. These are quite complex issues and though I give only an introductory gloss here, this subsection is, alas, somewhat hard going. However, unless we work through some issues about “group” selection and its relation to gene-culture and cultural evolution, it is impossible to understand Hayek’s views and why, I think, they are problematic. This subsection lays the foundation for important discussions in Part II about the basis of order in complex societies.

Multilevel Genetic Selection

Multilevel (or “group”) genetic selection is often subject to disputes as to precisely what constitutes a “group,” a complex matter that I shall leave aside.³⁶ For the evolution of trait α to be a case of group selection, what I shall call the “INDIVIDUALS” (lower-level subjects of selection) must in some sense qualify as “free-living” either now, or at least at some time in the past. These INDIVIDUALS must then interact in such a way that they form a stable unit (“the GROUP”) that is itself subject to selection pressures.³⁷ The basic logic of group selection is to distinguish the covariance between (i) trait α and the fitness of an INDIVIDUAL i and (ii) the covariance at the GROUP level between α and the fitness of GROUP G . The overall (total) fitness of trait α for individual i will depend on both factors. To see this better, say that α is the tendency of an individual i to act on reciprocity, and β a variant that is more selfish. The fitness of α *within* G for INDIVIDUAL i depends on the extent to which possessing α affects i 's fitness within G . We might suppose that INDIVIDUALS who are always ready to reciprocate and do their bit will be taken advantage of by the more selfish β types, so the fitness of α types will tend to be lower than β types within G . But suppose that an overall preponderance of selfish β renders GROUP G poorer because collective goods are not as well provided, and so having a preponderance of β types is associated with G being weak in GROUP conflicts, and so perhaps eliminated in conflicts by other GROUPS with a preponderance of reciprocity α types. So the overall fitness of the reciprocity trait depends on *both* its within-group INDIVIDUAL fitness and its between-group GROUP fitness. If the selection pressures between GROUPS are weak (group-level competition does not much contribute to fitness), we might expect β to dominate, as it seems more fit at the INDIVIDUAL level (selfish types usually go home richer than the reciprocators). On the other hand, if group-level pressures are great—if the main determinant of fitness is contribution to GROUP fitness, α (reciprocity), which is less fit at the INDIVIDUAL level, may still be, overall, more fit than β . Richer GROUPS (i.e., those with a preponderance of reciprocators) tend to eliminate poorer, β -preponderant GROUPS. Note that it is an error to ask whether α evolved through GROUP *or* INDIVIDUAL selection: the overall fitness of α depends on both, which is why “multilevel” selection is a more appropriate label.

In making a case that genetic group-level selection played an important role in the evolution of some trait α , it must be shown that (i) the GROUPS were sufficiently stable and distinct (cross-migration must be limited), (ii) there were significant genetic differences between the GROUPS, and (iii) group-level selection pressures for the fitness of α -types were strong, while at the individual level they were weaker. There has been protracted debate about whether in practice these have held in human evolution, and thus whether selection at the group level played an important role in genetic evolution. The strongest case for them to be met, I would venture, is gene-culture coevolution in

³⁶ See Sterelny, “The Return of the Group.” For an excellent overview, see Okasha, *Evolution and Levels of Selection*. In this very rough presentation, I follow the sophisticated analysis of Okasha. I avoid reference in this explication to the Price equation.

³⁷ There is no reason why there must be only two such levels; however, I shall only explore this simple case.

the Pleistocene era.³⁸ Because of the harsh climate in which human groups were often driven to the brink of extinction, the effectiveness of social cooperation was critical. And when there was conflict between groups for scarce resources, the effectiveness of the group as a cooperative unit would have been critical in determining success. Moreover, as we have seen (§11.1), culture can greatly speed up genetic evolution,³⁹ so that once humans were highly cultural beings, different cultural norms and practices could well have differentiated the genetic makeups of the competing groups. In this context, it has been argued, traits like altruism and strong reciprocity, which do not seem fitness enhancing at the individual level, were selected for.

Group Selection in Cultural Evolution

It is sometimes said that cultural group selection is a “variant” of genetic group selection models.⁴⁰ We must be clear about whether we are concerned with gene-culture evolution (which most definitely can be a variant of the “group” selection account of genetic changes) or—as is my concern here—cultural evolution per se. Whether an account of cultural evolution is a variant of multilevel selection, we can now see, depends on whether the evolution of some cultural trait γ is subject to selection at both a micro and macro level. If trait γ is subject to selection *only* at the group level, rather than a version of multilevel selection, what we really have is a straightforward “individual” selection where the INDIVIDUALS are cultural groups—there can then be no conflict between INDIVIDUAL- and GROUP-level selective pressures. To distinguish these from genuine multilevel accounts, let us call these “macro-level selection” analyses of cultural evolution.

At times Hayek suggests a macro-selection theory. He often stresses that, in cultural evolution, “the selection process of evolution will operate on the order as a whole”; what is selected, Hayek argues, is an “order of actions” that arises from numerous interacting rules, other elements of the social system, and the wider environment.⁴¹ This distinction between a set of underlying social rules and the order of actions or social order to which it gives rise is a fundamental insight of Hayek’s, which allows him to distinguish the focus of selective pressure (the order of actions) and the underlying rules that are transmitted from generation to generation. On Hayek’s analysis, a group living under a set of social rules R , composed of rules $\{R_1 \dots R_n\}$, will give rise to a certain abstract pattern of social interactions, O ; *it is O , the overall pattern, on which cultural selection operates*. A specific order O is an abstract pattern of a large number of human interactions (a “culture”), which does not arise from any specific rule R_i , or the aggregated effects of a set of independent rules, but from a set of interacting rules in an environment. Hayek described this as a sort of holism: “systems of rules of conduct will develop as a whole.”⁴²

³⁸ For the most powerful case, see Bowles and Gintis, *The Cooperative Species*. On the unstable climate of this era, see Richerson and Boyd, “Rethinking Paleoanthropology.” See also Henrich, *The Secret of Our Success*, pp. 207ff.

³⁹ Laland, Darwin’s *Unfinished Symphony*, pp. 216ff, 230ff.

⁴⁰ Tomasello, *A Natural History of Morality*, p. 12.

⁴¹ Hayek, “Notes on the Evolution of Systems of Rules of Conduct,” pp. 282–3.

⁴² *Ibid.*

Most readings of Hayek's analysis of cultural evolution understand it to be a form of macro selection at the social group level. As he says, "The rules of conduct have ... evolved because the groups who practiced them were more successful and displaced others."⁴³ However, this alone would not make the account a variant of multilevel selection: it simply has one level, the individual culture, which is in competition with other such individual cultures. Just as human phenotypes compete, and the more adaptive transmit their genes (which encode the information for building these phenotypes), on this view cultures compete, and the more adaptive transmit their rules (which encodes the information about how to build a successful culture). If society S_1 , characterized by order of actions O_1 , is more productive than S_2 based on O_2 , society S_1 will tend to win conflicts with S_2 , a mechanism akin to natural selection. Or, the members of S_2 , seeing the better-off participants in S_1 characterized by O_1 , may either emigrate to S_1 , or seek to copy the underlying rules R_1 , thus inducing differential rates of reproduction between the two sets of underlying rules.⁴⁴ It is important to stress that, to be a macro-level account, it is not enough to show that the rules of groups are replicated at differential rates (so that some rules tend to dominate more social groups): this can occur either because, at the individual level, these rules are more adaptive (and so individuals learn to employ them) or, as Hayek says, they are selected "because the groups who practiced them were more successful." That cultural rules evolve does not entail macro selection.

Selection that occurs only at the level of the "group" is not a variant of multilevel selection models in any interesting way. As we saw at the outset (§2.1), however, Hayek does recognize two levels of selection, providing the foundation for a genuine multilevel theory of cultural evolution.⁴⁵ His more nuanced view is that, while the "group level" is the primary locus of selection, rule selection also takes place in the form of competition between rules within a society—as he says, "cultural evolution operates largely through group selection."⁴⁶ For a rule R to be selected, it must be contributory to a selected order, O , but it must also attract allegiance within the group of individuals who coordinate via R . Individuals are constantly testing rules to determine whether conformity suits their overall concerns; "it is, in fact, desirable that the rules should be observed only in most instances and that the individual should be able to transgress them when it seems to him worthwhile to incur the odium this will cause.... It is this flexibility of voluntary rules which in the field of morals makes gradual evolution and spontaneous growth possible, which allows further modifications and

⁴³ F. A. Hayek, *Rules and Order*, p. 18; Hayek, *The Fatal Conceit*, p. 25.

⁴⁴ Hayek, *The Political Order of a Free People*, p. 159; Hayek, *Rules and Order*, p. 47; Hayek, *The Fatal Conceit*, pp. 6, 25, 43. But see §23.4 in this volume.

⁴⁵ Indeed, he also advances an account of gene-culture evolution, in which the success of a group affects the selection of individual genetic traits within it, allowing traits that have an in-group disadvantage to be selected. "Although the existence and preservation of the order of actions of a group can be accounted for only from the rules of conduct which individuals obey, these rules of conduct have developed because the individuals have been living in groups whose structures have gradually changed. In other words, the properties of the individuals which are significant for the existence and preservation of the group, and through this also for the existence and preservation of the individuals themselves, have been shaped by the selection of those individuals from the individuals living in groups which at each stage of evolution of the group tended to act according to such rules as made the group more efficient." Hayek, "Notes on the Evolution of Systems of Rules of Conduct," pp. 283–4.

⁴⁶ *The Fatal Conceit*, p. 23, emphasis added.

improvements.”⁴⁷ Although Hayek himself disparaged rule selection based on how well a rule conformed to one’s social or moral ideals,⁴⁸ we have seen that individual sensitivity to a norm is critical for its stability, and for that it needs to align with an INDIVIDUAL’s normative commitments (§9.3). One of the factors that determine within-group fitness of a moral rule is thus its ability to secure such allegiance and be taught to the next generation.⁴⁹ This is a case of what Richerson and Boyd call “content bias”:⁵⁰ rules that accord with people’s moral sensibilities are more apt to be learned and transmitted.

On this analysis, the evolution of a cultural trait γ depends on its fitness within the GROUP (the extent to which individuals adopt it) and its contribution to GROUP fitness (in competition with other orders of actions). We can now apply our lessons from our discussion of multilevel selection: if there are two levels of selection operating on a moral rule, then the strength of the different selective pressures will be critical in determining which of the two is more influential.⁵¹ If cultural competition decreases or moderates, then we can expect INDIVIDUAL-level selection to be a much stronger determinant. An interesting hypothesis thus emerges: if cultural evolution is genuinely multilevel in this way, then in eras of decreased GROUP competition, public justification of a rule should be a much stronger force in a rule’s selection as it is competing for individual agents’ endorsement against alternative rules. On the other hand, in eras of intense inter-group competition, we should expect that rules that are not well-aligned with the moral commitments of individuals, but which are selected at the group level, may predominate.⁵²

This, in turn, leads to another important point. In multilevel analysis, effective higher-level GROUP selection inherently restrains lower-level INDIVIDUAL selection. There really is no point to invoking higher-level selection if it does not. In the evolution of cooperation literature, the point of invoking a GROUP-level selection is to restrain the success of INDIVIDUAL (selfish) agents so that within-group less adaptive, cooperative agents can thrive. A mammal can be seen as a case of GROUP-level selection, insofar as the possible strategies of individual cells are constrained by the adaptive needs of the GROUP (individual mammal). A cancer cell is precisely a part that has broken free of these restraints, and because of this threatens ultimate system collapse. We might say, in a rough and ready way, that restricting the social influence

⁴⁷ Hayek, *The Constitution of Liberty*, p. 63

⁴⁸ On page 161 of the Epilogue of *The Political Order of a Free Society*, Hayek argues that the steps in cultural evolution toward large-scale coordination “were made possible by some individuals breaking some traditional rules and practising new forms of conduct—not because they understood them to be better, but because the groups which acted on them prospered more and grew.” For a general analysis of the role of conscious deliberation and choice of rules in Hayek, see Pert and Levy, “Discussion, Construction and Evolution.”

⁴⁹ On the importance of teaching in human cultural evolution, see Laland, *Darwin’s Unfinished Symphony*, esp. pp. 156ff.

⁵⁰ Richerson and Boyd, *Not by Genes Alone*, chap. 3.

⁵¹ “The theory of cultural multilevel selection says that this [higher-level] evolution is only possible when societies compete against each other, so that those lacking the right institutions fail.” Turchin, *Ultrasociety*, p. 19.

⁵² Note that this is not an implication of a simple “macro”-selection account, which only posits selection pressures at the group level. Thus the importance of distinguishing genuine multilevel accounts for simple “group selection.”

of INDIVIDUAL-level preferences (in which a rule's within-group fitness is determined by its attractiveness to individuals) in order to secure system-wide functionality is precisely what GROUP selection accomplishes. If GROUP-level pressures are great, the rules will be less responsive to the aims of the INDIVIDUAL agents, and indeed significantly restricting their actions will be critical to the culture's success. When GROUP-level selection is strong, it is entirely appropriate to call a culture a "superorganism."⁵³ In such a culture, rules will tend to be more restrictive, and public justification may be less important.

§12 The Rise and (Partial) Fall of Inequality

12.1. Clans

In the previous section we have seen that a natural history of the evolution of human morality also is a history of the evolution of culture. Too often the "evolution of morality" is seen as a story that ends around 40,000 years ago, since it was often thought that the genetic evolution story pretty much ends there. But since cultural evolution has tremendously speeded up, and indeed now swamps both genetic and gene-culture evolution, the natural history of humans as a normative species converges with their evolution as cultural beings. Consequently, we have been focused on the ways humans learn from each other, and how moral rules culturally evolve. The natural history of humans as a normative species by no means concludes with Late Pleistocene egalitarianism. Let us pick up our historical narrative, now that we have a better understanding of morality as a manifestation of culture as well as nature.

In a familiar account, an inegalitarian counter-revolution set in quickly with the emergence of agriculture in the Holocene era, starting around 10,000–12,000 years ago. More recent research pushes the origins of inequality further back into the end of the last ice age, say 17,000 years ago.¹ A critical development in the rise of inequality was the development of forager bands with clan-based organization.² The development of societies based on lineages, clans, and subclans was more formal and hierarchical than clanless hunter-gatherer groups, and achieved considerably higher population density.³ Clans are typically symbolic and mythical family lineages; while there is often a claim of common descent, one is not born into a clan but must be initiated—a culturally complicated and often painful form of joining.⁴ For our purposes, what is critical is that clans seek to extend social organization beyond close kin and band affiliation by employing cultural symbols to mark this extended conception of kinship. This expanded social organization allowed for larger organized

⁵³ Henrich, *The Secret of Our Success*, p. 318.

¹ Flannery and Marcus, *The Creation of Inequality*, p. 15.

² Weber is distinctive in analyzing the role of clans in economic history. *General Economic History*, Part I.

³ Flannery and Marcus, *The Creation of Inequality*, p. 16.

⁴ *Ibid.*, pp. 16, 50, 55.

populations, which often provided decisive advantages in conflict over small-band foragers and in food-sharing.⁵

Whereas close kin do not, as it were, have to work to create kinship ties, these extended, essentially artificial, kinship units devoted extensive symbolic and cultural resources to marking themselves off as a common unit. It is here, perhaps, that the distinction between “us” and “them” was intensively cultivated through cultural forms, and which provided a decisive cultural context for the evolution of our so-called groupish or tribal sentiments and mindset.⁶ It is well-known that humans easily divide into groups,⁷ and that we often think of intention and plans in the first-person plural “we,” as well as the singular “I.” Some accounts place these developments very early in human gene-culture coevolution—in Tomasello’s view, human sociality and moral life are essentially a tale about the evolution of collective intentionality and thus early on presuppose collective agents.⁸ Now without denying that this is part of an adequate account, there seems reason to suppose that a critical phase of growth of these groupish sentiments occurred much later, and was a mechanism *not* for the *constriction of social relations*, but for deepening commitments with a *wider* group of unrelated people. This deserves emphasis: the relatively common view that these sentiments are an inheritance of our hunter-gatherer egalitarian past and its small “closed” society is misleading (§1.1):⁹ a highly plausible hypothesis is that “groupishness” was an evolutionary mechanism to *expand* social groupings further than close kin-based and forager-sized band cooperation. We must remember E. O. Wilson’s observation that nepotism resulting from kin selection is “the enemy of civilization.”¹⁰ Clan civilization’s expansion of loyalties far beyond close kin required powerful symbolic and psychological mechanisms (including often excruciating clan initiation rituals) to develop an identity as “us,” to be distinguished from those outside the clan. A mechanism for forming a more intense common identity may well have built on the already well-developed tendency to (*i*) be sensitive to ethnic markers and (*ii*) conform to majority behavior—we have already seen that humans excel in their ability to imitate (§11). As Over and Carpenter argue, “over-imitation” signals group membership: “One of the most important messages that imitation may convey . . . is ‘I am like you’ or, at a group level, ‘I am one of you.’”¹¹ Of course, as we saw (§6.4), *Pan* probably possessed some tendencies toward “xenophobia,” so some preadaptation for such “groupishness” was likely present in humans.

Unsurprisingly, once the collective identities are formed, the “other” could easily become the enemy: archaeological evidence indicates that clan-based societies were among the earliest perpetrators of intense group-based violence.¹² The relation between pro-group (e.g., altruistic and trusting) behavior and violence has been

⁵ Ibid., pp. 16–7.

⁶ Ibid., p. 40.

⁷ The classic study here is Tajfel et al., “Social Categorization and Intergroup Behavior.” See also Greene, *Moral Tribes*, pp. 48–55.

⁸ This is the theme of his *Natural History of Morality*.

⁹ It is also misleading about hunter-gatherer societies which, we have seen, can be relatively open to social networks with neighboring bands, and so are not obviously “closed.”

¹⁰ Wilson, *On Human Nature*, p. 157.

¹¹ Over and Carpenter, “The Social Side of Imitation,” p. 8.

¹² Flannery and Marcus, *The Creation of Inequality*, pp. 40ff.

supported by experiments among populations recently subject to extreme violence. In experiments conducted by Voors in Burundi after the horrific civil war, it was found that those subjected to violent conflict displayed significantly increased levels of pro-social motivations.¹³ This provides evidence for the familiar claim that in times of conflict, humans think more in terms of “us” than “me.” Gilligany, Pasquale, and Samii, in their study of communities exposed to violence during the civil war in Nepal, also found increased tendency to contribute in public goods games and were significantly more trusting of others—in general, we can say social cohesion increased. However, Gilligany, Pasquale, and Samii argue that the more powerful explanation is not a transformation to more groupish preferences, but a change in social norms and institutions.¹⁴ This is an important result. Although it is highly plausible that conflict produces more group-oriented preferences,¹⁵ groups can quickly adjust to the need for intensified cooperative behavior by altering norms and institutions. Perhaps fidelity to these norms is supported by increased pro-social motivations, but our capacity for norm-guided behavior (§9.2) seems a critical, perhaps the most important, variable in explaining increased social cohesion in conflictual periods. Even in these extreme cases, our ability to be norm guided may well be at least as important as our tendency to become “tribal.”

As Flannery and Marcus suggest, in a Rousseauian-inspired history of inequality, clan-based societies can be seen as the beginning of a hierarchical-inegalitarian counterrevolution: already in such non-agricultural societies we find the hierarchies in social status that Rousseau thought decisive in the rise of inequality.¹⁶ While there was, of course, great variance throughout the world, two generalizations appear reasonable. *First*, evidence indicates that as clan-based societies arose, there was explicit effort by some to attain positions of prestige.¹⁷ In the simple tale in which hunter-gatherers were thoroughly and “instinctively” egalitarian, it is a mystery how inequality began to rise so quickly in the last ice age, and then spectacularly around 10,000 years ago. Recall, however, that on Boehm’s widely accepted account, egalitarianism was a “reverse dominance hierarchy” (§7.4) in which subordinates explicitly policed potential alphas (such as Cephu) seeking to play the role of “big men” or “chiefs.” The human tendency, so strong in *Pan*, for hierarchy and dominance did not go away: it was well controlled, but always ready to reassert itself given the opportunity. However, Peter Turchin rightly points out that it is inadequate to claim that the new clan leaders or priests simply hoodwinked or bullied subordinates into recognizing them as prestigious leaders. Egalitarian societies had been remarkably successful for a very long period—perhaps a hundred thousand years or more—in keeping would-be bullies under control.¹⁸ A second factor seems necessary: with the expansion of group size induced by clans, hierarchy became more functional—and egalitarianism had great

¹³ Voors et al., “Violent Conflict and Behavior: A Field Experiment in Burundi.”

¹⁴ Gilligany, Pasquale, and Samii, “Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal”; “Civil War and Social Capital: Behavioral-Game Evidence from Nepal.”

¹⁵ Henrich, *The Secret of Our Success*, pp. 206ff; Haidt, *The Righteous Mind*, pp. 162ff.

¹⁶ Flannery and Marcus, *The Creation of Inequality*, p. 86.

¹⁷ *Ibid.*, p. 191.

¹⁸ Turchin, *Ultrasociety*, pp. 143ff.

difficulties scaling up.¹⁹ We have sound evidence that when conflict occurs, group size is a critical factor in determining success.²⁰ If hierarchical norms allowed the effective governance of larger and more successful groups, those populations that tolerated its re-emergence would tend to thrive over their more egalitarian neighbors. Here multi-level cultural selection may well have been decisive: norms that were relatively poorly aligned with the normative sentiments of the mass of citizens (§9.3) were highly functional in larger groups, especially those with high inter-group conflicts.

Clan-based societies certainly displayed inequality both within the clan and, in larger societies with multiple clans, between clans. Within clans, those who claimed closer descent from the Common Ancestor often also claimed social prestige, as did those who were adept at the rituals on which clan unity was based. Gradations of prestige in clan societies sometime became stratified into distinct classes in “rank societies,” in which differential status became hereditary. On Flannery and Marcus’s reading of the data, there was often cycling within a single society, as ranks were established, then challenged by a resurgent egalitarianism.²¹

12.2. Agriculture and the State

On the traditional account, the discovery of agriculture gave rise to the need for a military to protect cultivated land and a now immobile population, and that in turn engendered the early archaic state and its extreme inequality. As usual, the closer we look, the more complex the story is. For one, sedentary populations predate agriculture.²² And there were egalitarian agricultural communities, while the rise of agriculture predated the state, in some instances, by thousands of years.²³ Nevertheless, as Rousseau held, agriculture and the accompanying individual- and family-based claims to own land and its produce provide the foundation for high wealth inequalities.²⁴ But it is the rise of the early state—say between 4,000 and 3,000 BCE in Mesopotamia—that essentially did away with equality (and indeed freedom), replacing them with a harsh, highly coercive, hierarchical order.²⁵ In Sumer, one of the earliest states in Mesopotamia, “Hundreds of rules of social behavior had allegedly been established by the gods; human priests, judges, and bailiffs were there to make sure that they were followed. The state decided what men were allowed to do, what women were allowed to do, who could marry, who could divorce, who could strike

¹⁹ It has been argued that status hierarchies assist in the provision of public goods, a critical problem in larger groups. See Mark, “Status Organizes Cooperation: An Evolutionary Theory of Status and Social Order.”

²⁰ See Richerson and Boyd’s discussion of the conflict between the Nuer and the Dinka. *Not by Genes Alone*, pp. 23–6.

²¹ Flannery and Marcus, *The Creation of Inequality*, pp. 198ff.

²² Scott, *Against the Grain*, pp. 10, 49ff.

²³ Turchin, *Ultrasociety*, pp. 91, 136ff, 208. See also Scott, *Against the Grain*.

²⁴ “The first man who, having enclosed a piece of land, to whom it occurred to say *this is mine*, and found people sufficiently simple to believe him, was the true founder of civil society.” Rousseau, *The Discourse on Inequality*, p. 164. See Turchin, *Ultrasociety*, p. 175.

²⁵ Scott, *Against the Grain*, p. 157. “Nobles in archaic states had many more rights than commoners, while commoners were weighed down with obligations and slavery was common. At the summit of the social hierarchy, a ruler could be ‘deified’—treated as a living god.” Turchin, *Ultrasociety*, pp. 132–3.

whom, and so on. The penalties included fines, corporal punishment, and even death by stoning.”²⁶

In his recent groundbreaking study, James C. Scott argues that grain-based monoculture, rather than being a cause of this new, deeply hierarchical state, was essentially its creation.

The key to the nexus between grains and states lies, I believe, in the fact that only the cereal grains can serve as a basis for taxation: visible, divisible, assessable, storable, transportable, and “rationable.” Other crops—legumes, tubers, and starch plants—have some of these desirable state-adapted qualities, but none has all of these advantages. To appreciate the unique advantages of the cereal grains, it helps to place yourself in the sandals of an ancient tax-collection official interested, above all, in the ease and efficiency of appropriation. The fact that cereal grains grow above ground and ripen at roughly the same time makes the job of any would-be taxman that much easier.²⁷

Thus, rather than concluding that “something strange happened to human societies when they adopted agriculture,”²⁸ we perhaps should conclude that something strange happened when we adopted the state. Rather than seeing the hierarchical state as the inevitable result of agrarian life, it now looks as if many of the features of agrarian life were—if not its creation—certainly intensified by it. Again, this is not to say that the rise of the state was nothing more than a plot by alphas to regain control—the state’s organization gave it decisive military advantages over more egalitarian groups²⁹—but it also exercised great influence over its own environment, reinforcing inequality through the imposition of taxable ways of life.

12.3. A U-Shaped Inequality Curve? A Backward-J Equality Curve?

The developments we have thus far canvassed have been depicted by a U-shaped inequality curve: strong hierarchy and despotism in *Pan*, with a nadir of inequality and hierarchy among LPA hunter-gatherers, returning to extreme degrees of inequality in the archaic states.³⁰ This curve must mystify those who think humans are either essentially domineering Machiavellians or instinctive egalitarian sharers. We neither left hierarchy and domination behind once and for all when evolving Late Pleistocene

²⁶ Flannery and Marcus, *The Creation of Inequality*, p. 477.

²⁷ Scott, *Against the Grain*, pp. 129–30.

²⁸ Turchin, *Ultrasociety*, p. 135.

²⁹ “Groups of poorly nourished—perhaps even chronically sick—farmers were able to exterminate healthy and tall foragers simply by force of numbers.” Turchin, *Ultrasociety*, p. 174. Turchin analyzes the advantages of greater numbers in detail. For a discussion of Turchin’s work, see Mesoudi, *Cultural Evolution*, pp. 125ff.

³⁰ Boehm, *Hierarchy in the Forest*, p. 65. Cf. Turchin: “In their degree of despotism, the archaic states far exceeded even the ancestral great-ape societies. Chimps and gorillas do not deify their alpha males, perform ape sacrifice, or enslave other members of the troop.” *Ultrasociety*, p. 180.

egalitarian societies, nor are we consistently domineering Machiavellians. Boehm's analysis is, once again, compelling:

the same quite definite and "hierarchical" human political nature could have been supporting not only despotic societies of recent humans and ancestral apes, but also the egalitarian societies of humans. In despotic social dominance hierarchies the pyramid of power is pointed upward, with one or a few individuals (usually male) at the top exerting authority over a submissive rank and file. In egalitarian hierarchies the pyramid of power is turned upside down, with a politically united rank and file decisively dominating the alpha-male types.³¹

Of course, along the way, there were critical developments—what I have called the Modern Egalitarian Moral Package—including the evolution of normative guidance, reciprocity, and social learning. The germs of subordinate rebellion in *Pan* (§6.3) evolved into a much stronger concern with fairness and aversion to being dominated by bullies. Nevertheless, that humans started out deeply hierarchical, became radically egalitarian, only to revert to extreme hierarchy, should make us wary of claims that we are "by nature" fit for one sort of society or another. At the same time, it should make us suspicious of those who hold that we are so plastic we can fit into almost any social order. Late Pleistocene egalitarianism was achieved through a social structure that effectively channeled strong human propensities: an egalitarian morality that regulated holdings and prestige, thwarting bullies but also restraining progressive individuals who would seek to outproduce others. Human social life has always been a complex interplay of nature and norms, cooperative regulation and individual aims. We are constantly tempted to stress one at the expense of the other, rendering social theory either unduly conservative and pessimistic about social change, or unrealistically optimistic that social structure can produce any social relations our ideals valorize.

Still, the U-shaped *inequality* curve seems too pessimistic. During our long egalitarian era, humans evolved sentiments that strongly incline most of us to egalitarian norms and systems of cooperation based on reciprocity. We certainly did not end up where we started. Moreover, we should not make the mistake of conceiving all hierarchy as a form of domination: *Pan* and archaic state hierarchy certainly was, but modern life is a mosaic of hierarchies that oppress and ones that help all secure their purposes.³² An alternative view is that we have traversed a backward-J *equality* curve since Late Pleistocene societies: starting out with very high equality, with a nadir in the archaic states, and a general direction of increased equality since, though by no means returning to the high levels of Late Pleistocene societies. This is Turchin's view.³³ Following Robert Bellah, Turchin argues that an important move back toward equality was made with the development of large multinational empires. Whereas some archaic states had populations in the hundreds of thousands, the Persian Empire had a population of 30 million, Rome something like 60 million. The hypothesis is that the need to find norm endorsement in such huge diverse populations pushed these empires toward more impartial and tolerant religions and norms—"a general

³¹ Boehm, *Hierarchy in the Forest*, pp. 63–4.

³² The lesson of Coase, "The Nature of the Firm."

³³ Turchin, *Ultrasociety*, pp. 22, 132.

morality.”³⁴ Note that this is the implication of our model in section 10.3: huge multi-ethnic empires needed to “climb the ladder” of impartiality as they grew if their norms and laws were to achieve the required level of justification to ensure stability. Breaking out of the confines of ethnically based states and religions, these empires engendered the first germs of a cosmopolitan citizenship and a common human standing.³⁵ It also should be stressed that, while the archaic states were generally ethnically based, and so could be considered as instances of human “groupish” cultural units, the swift appearance of stable empires itself indicates that new norms can engender cooperation across diverse ethnic groups. Cultural evolution and social learning can proceed quickly, and in ways that drastically relax the inward pull of group membership.

Again, we must beware the temptation of romanticizing hunter-gatherer morality. While it may have been the high point of equality on many dimensions, Late Pleistocene egalitarian life manifested inequalities that we would no longer tolerate, such as in gender roles (§7.4). And it was certainly not the apex of freedom—though it certainly was not the nadir, as might be suggested by some accounts of “closed tribal societies.” (The archaic societies seem to have the dubious distinction of being the nadir of both freedom and equality—until the twentieth century). Autonomy of subordinates was protected against alpha male domination, but at the costs involved by reverse domination, which could bring high levels of coercion to bear on those who sought improvements for themselves. Interestingly, whereas Boehm, the ethnographer, stresses the way in which charges of witchcraft are brought against would-be dominators and those who are stingy or too aggressive,³⁶ Platteau, the developmental economist, stresses the way in which such charges are used against those who show themselves to be most productive and progressive. Both are correct.

12.4. The Rise of WEIRD Morality

In an important study Henrich, Heine, and Norenzayan argued that on a wide variety of psychological measures—including reasoning styles, fairness, cooperation, spatial reasoning, moral reasoning, and self-concepts—those in Western Educated Industrialized Rich Democracies (WEIRD societies) were outliers, often extreme outliers, in experiments.³⁷ A number of these measures that concerned morality were pioneered by Jonathan Haidt. In his 2012 book *The Righteous Mind*, Haidt developed this idea of WEIRD morality in depth, contrasting it to moralities upheld by most people around the world. Roughly, Haidt’s data indicates that moral reasoning of WEIRD populations (especially the university educated) is largely focused on individuals (rather than groups or social relations),³⁸ and centers on the dimensions of

³⁴ Many have stressed the importance of the “great religions” in promoting “universal morality.” For a general discussion in the context of moving beyond small-scale morality, see Platteau, *Institutions, Social Norms and Economic Development*, pp. 256ff.

³⁵ Turchin, *Ultrasociety*, chap. 9.

³⁶ Boehm, *Hierarchy in the Forest*, p. 84; Platteau, *Institutions, Social Norms and Economic Development*, p. 201.

³⁷ Henrich, Heine, and Norenzayan, “The Weirdest People in the World?”

³⁸ This individualism is confirmed by other studies. *Ibid.*

liberty–oppression, care–harm, and fairness–cheating. In contrast, most other moral systems (including, Haidt argued, conservatives and some minorities in WEIRD societies) include, *in addition*, loyalty–betrayal, authority–subversion, and sanctity–degradation dimensions of moral concern.³⁹ To simplify even more, the core idea is that the morality of liberalism is a morality about helping, harming, freedom, oppression, fairness, and cheating. If one takes as canonical liberal texts Mill’s *On Liberty* and Rawls’s *Theory of Justice*, one would get the essentials of WEIRD morality. It is a cosmopolitan morality of impartiality, and typically sees moral progress in terms of increasingly impartial, fair relations between free and equal individuals.⁴⁰

In a recent wide-ranging and detailed study, Schulz, Barahmi-Rad, Beauchamp, and Henrich have sought to discover the origins of this WEIRD way of thinking.⁴¹ Drawing on a wide range of anthropological and psychological studies, they provide impressive evidence that a key variable was the breakup of kin- and clan-based morality by the Catholic Church over a long span of time. Most moral systems are focused on extensive kinship networks—as we have seen, critical to expanding the sphere of human cooperation was to expand the concept of kin to include the clan (§12.1) and beyond. As they observe, “Intensive kinship norms reward greater conformity, obedience, holistic/relational awareness and in-group loyalty but discourage individualism, independence and analytical thinking. Since the sociality of intensive kinship is based on people’s interpersonal embeddedness, adapting to these institutions tends to reduce people’s inclinations towards impartiality, universal (non-relational) moral principles and impersonal trust, fairness and cooperation.”⁴² Kin- and clan-based moralities maintain their ingroup commitments by allowing, sometimes requiring, marriages between types of cousins and other relatives. The details of these marriage norms are not crucial here: what is critical is that in some way marriage within the kin or clan network is encouraged, which solidifies its boundaries and mutual obligations. It was a unique feature of the Roman Catholic Church that it has an almost obsessive concern with discouraging incest very broadly understood:

the Church’s Marriage and Family Program (MFP) began with targeted bans on certain marriage practices used to sustain marriage alliances between families (e.g., levirate marriage); however, by the Early Middle Ages, the Church had become obsessed with incest and had begun to expand the circle of forbidden relatives, eventually including not only distant cousins but also step-relatives, in-laws and spiritual-kin. Early in the new millennium, the ban was stretched out to encompass sixth cousins, including all affines. . . . At the same time, the Church promoted marriage “by choice” (no arranged marriages) and often required newly married couples to set up independent households (neolocal residence). The Church also forced an end to many lineages by eliminating legal adoption, remarriage and all forms of polygamous marriage as well as concubinage, which meant that many lineages began literally dying out as they lacked legitimate heirs.⁴³

³⁹ Haidt, *The Righteous Mind*, pp. 181–2.

⁴⁰ See, e.g., Kitcher, *The Ethical Project*; Buchanan and Powell, *The Evolution of Moral Progress*.

⁴¹ Schulz, Barahmi-Rad, Beauchamp, and Henrich, “The Origins of WEIRD Psychology.”

⁴² *Ibid.*, p. 2; citations deleted.

⁴³ *Ibid.*

Schulz, Barahmi-Rad, Beauchamp, and Henrich provide extensive evidence that a critical predictor of a population's "WEIRD-ness" is the length of time spent under the Catholic Church. Extending their analysis to those parts of the world settled by European immigrants, evidence indicates that the thinking of adult children of immigrants is significantly predicted by their parents' native countries or their originating ethno-linguistic populations.

As kin-based moralities had been undermined for over a thousand years, Schulz, Barahmi-Rad, Beauchamp, and Henrich argue that the resulting morality among essentially unrelated individuals stressed, among other traits, familiar WEIRD features such as individual choice, impartiality, fairness with strangers, and outgroup trust, while discouraging nepotism and particularism. On the basis of their study, it seems that there has indeed been a real divide between kin- and clan-based moralities and the individualistic morality of impartiality. This is not a conflict between an "atavistic" "tribal" human nature left over from Late Pleistocene society and the Open Society, but between the dominant heavily kin-based normative systems and the individualistic impartial morality of WEIRD societies. Again, we see that "human nature" can be guided by very different norm systems, with drastically different consequences.

It seems that WEIRD societies were forced to scale up moral relations in a unique way. Rather than developing and expanding clan-based relations, they were ultimately forced to build morality on an individualistic basis, stressing impartiality, rather than the more ethnocentric focus of kin-based systems. In this sense they had a deprived basis from which to work: rather than appealing to notions of sanctity, purity, and hierarchy, they were left to reconstruct moral relations on the basis of essentially nuclear families who, compared to those embedded in kin networks and obligations, were free and equal. Thus individual-centered concerns—liberty, absence from being bullied and harmed, fairness in dealing with unrelated individuals—dominated. But this very privation was conducive to WEIRD morality's impartiality. It could "climb the ladder" of impartiality (critically, which is its own understanding of moral progress) because it had to develop without relying on appeal to particularism and nepotism. As many of its critics have claimed, liberal morality was indeed premised on (the liberal WEIRD idea of) abstract individuals (who think in terms of other abstract individuals) rather than socially embedded kin systems. Just because of this, however, it has the ability to scale up indefinitely to encompass ever-larger populations of those who—if not exclusively, dominantly—conceive of themselves as free and equal individuals.

§13 A Complex Moral Species

13.1. An Ambivalent Species

The history of evolutionary moral and social theorizing has, regrettably, too often depicted us as a rather simple species that has a true nature, though it is often masked by social forms. This was the great weaknesses of Hayek's belief that we have atavistic

egalitarian instincts that the Great Society valiantly endeavors to control.¹ Evolution did not begin in the Late Pleistocene period. Hayek was as wrong as Rousseau about our egalitarian origins. But neither is our nature set by *Pan*: it is equally false that you can “scratch an ‘altruist’ and watch a ‘hypocrite’ bleed.”² We are no more Machiavellian strategists than egalitarians—and perhaps no less.

This Part has shown us to be an ambivalent and complex species, with tendencies running in different, sometimes opposite, directions.³ Most of us are adamantly concerned with freedom from domination, yet quickly adapt to hierarchies, some of which are inherently dominating, while others are mutually beneficial and publicly justifiable. Some of us are always on the lookout to dominate others and are tempted to exploit all hierarchy as a means to dominance. But our efforts to control domination themselves all too often end up dominating progressive individuals. We are self-interested strategists, yet one of our critical capacities is to copy those around us, often the most prestigious—which leads us back to the conflict between submitting to and resisting hierarchy. Our ability to be normatively guided allows us to expand our networks of cooperation, but this ability itself employs ethnic markers that restrict our cooperation. Our consciences are critical in allowing us to be guided by shared norms, but these very consciences are “flexible,” sensitive to possibilities to find ways to promote self- and kin-interests. Indeed, our very idea of what constitutes morality tends to be informed by our interests, and in moral justification we often act more like advocates than impartial judges. Our much discussed groupish “tribal” sentiments, which seem so exclusivist, may well have been encouraged as a way to expand cooperation beyond close kin to clan and then to tribes and ethnic groups.

While evolutionary theorists have often succumbed to an overly unified and simple picture of humans, they are nevertheless entirely correct that the possibilities of humans are limited by our complex evolved nature, and the culture that has evolved along with us and has in turn shaped us. For example, given the evidence, it seems most unlikely that social orders can persist that require unanimity of belief about the good or notions of the right. Humans inherently disagree about these matters: methods of reconciliation are part and parcel of humans living together, from (at least) Late Pleistocene societies on. Extreme coercion—what Rawls called “the oppressive use of state power”⁴—can serve as a way of “reconciliation” by subjecting all to the views of those at the top of the hierarchy, but even this does not suppose a grounding of social life in a homogenous unanimity. Neither can a viable social order be grounded in persistent extreme devotion to the good of the group at great costs to self—though in periods of extreme conflict, even such orders may be viable (§12.1). Just as we cannot stably immerse ourselves in a commonweal, neither can we exist in social life where social relations with non-kin are based purely on Machiavellian self-interest. The assumption in so much of economics and philosophy that starting with pure egoistic agents is somehow a neutral basis for modeling social life is manifestly false, as Rousseau so clearly saw. Reciprocity—which, I have stressed, is not a form of

¹ Hayek, *The Political Order of a Free People*, p. 160.

² Ghiselin, *The Economy of Nature and the Evolution of Sex*, p. 247.

³ Chapman, “Toward a General Theory of Human Nature and Dynamics.”

⁴ Rawls, *Political Liberalism*, p. 37.

tit-for-tat egoism—is a core feature of human social relations, and not simply personal relations. While, of course, we are capable of feats of altruism (especially toward kin), the basic moral norms of viable cooperative orders must be based on relations of reciprocity, as Smith realized about the market (§8.3).

13.2. A Norm-Guided Species

Still, the space of human social possibilities is large. Our very ambivalences and tensions allow us to sustain an amazingly wide range of social orders. Being, as it were, not optimized to any environment, we can cope with a wide variety of them. Every type of order differentially encourages some aspects of our nature while discouraging others, from hierarchy to egalitarianism, from narrow cooperation to wide cooperation, from impartial WEIRD morality to kin-based ethics.

In this Part, however, I have stressed the importance of normative guidance—the ability to coordinate on shared rules, internalize them, and hold others accountable to them. Many of our so-called tribal conflicts are about what norms we should adopt: as I have stressed, morality is often about choosing sides (§10). Our ability to adopt and adapt to a wide variety of norms renders humans a distinctly moral species: we achieved ultra-sociality with non-kin via the guidance of moral rules and norms. Even phenomena that at first inspection seem obviously based on “tribal sentiments,” such as intense devotion to pro-social aims during times of inter-group conflict, seem critically affected by changes in the relevant norms (§12.1). Norms can adjust to vast ranges of social circumstances; as norms adjust, we can swiftly adjust our social relations. For too long, social and political theorists have held that specific sentiments (such as egalitarianism or aggression) ground types of societies. Although such sentiments limit the range of the feasible, within that vast range our ability to be guided by shared social rules allows for a great diversity of viable social orders.

What, though, of Late Pleistocene “egalitarian sentiments” in the sense of a natural tendency to prefer equal distributions? Surely, it might be said, here Hayek was right that these substantive sharing “intuitions” were critical to Late Pleistocene life, and remain a strong pull on us today. In the last twenty years there has been a great deal of attention paid to so-called inequity aversion in humans. Ernst Fehr and Klaus M. Schmidt proposed an inequity aversion model which, they argued, accounted for people’s choices in experiments on the Ultimatum Game and other bargaining and sharing games.⁵ An Ultimatum Game involves two subjects, Proposer and Responder, who have X amount of money to distribute between them. In the simplest version of the game, Proposer makes the first move, and gives an offer of the form, “I will take n , leaving you with $X-n$,” where n is not greater than X . If Responder accepts, each gets what Proposer offers; if Responder rejects, each receives nothing. If players cared only about the amount of X that they received and were instrumentally rational, it would be rational for Proposer to, say, take 9 out of 10, offering Responder 1. As the second decider, Responder would be faced with a choice between 1 and nothing, and since an instrumentally rational agent must choose more over less, she will accept the

⁵ Fehr and Schmidt, “A Theory of Fairness, Competition, and Cooperation.”

offer. Since Proposer knows this, and since Proposer also will not choose less over more, Proposer will make that offer. This is not the observed outcome. In the United States and many other countries, one-shot ultimatum games result in median offers of Proposers to Responders of between 50% and 40%, with mean offers being 30% to 40%. Responders refuse offers of less than 20% about half the time. Fehr and Schmidt explained such results by proposing a utility function in which individuals have an aversion to inequity in outcomes, thus perhaps supporting the view that we are “naturally” inclined toward equal distributions.

Although the Fehr and Schmidt model has been widely influential, it is highly controversial, both methodologically⁶ and substantively. Substantively, Bicchieri has argued that evidence supports an alternative hypothesis: that people’s fairness reactions in these games are mediated by what choices the Proposer had. In her view, this indicates that perceived norm violation is critical: Did the Proposer make counter-norm choices, or was the offer consistent with the choices permitted by norms? The same offer thus may be accepted or rejected depending on whether the Responder thinks the Proposer’s choice violated a fairness norm.⁷ This general conclusion has been supported by López-Pérez, who argues that

a large body of experimental evidence, including very different phenomena like generous and punishing behavior, may be explained by a relatively simple utility theory in which agents experience different emotional responses depending on how they and others act. Roughly speaking, our claim is that aggressive emotions like anger and moral emotions like guilt or shame are strong psychological forces that enforce reciprocity, understanding by that concept two things: (1) people adhere to norms if they expect others to respect them as well, and (2) people punish those who violate binding norms. Further, and since these emotions are activated by deviations from the norm, they induce path-dependent preferences.⁸

Evidence is increasing that subjects’ decisions are not simply determined by comparative outcomes or shares, but by the choices that (for example) Proposers make on the way to the outcomes.⁹ As López-Pérez indicates, a good deal of evidence supports a *reactive emotions theory*: Responders experience emotional reactions such as anger, irritation, indignation, and contempt, based on their perceptions of fair or unfair choices.¹⁰ This path-dependence of reactions points to the relevance of norms of fairness: Did the other player violate norms of fairness on her way to the offer?

The norm-based explanation of fairness responses in the Ultimatum Game and related games is greatly enhanced by cross-cultural research: play in small-scale societies can radically differ from the usual result observed in market societies. And it is most certainly not the case that small-scale societies are more apt to play the games

⁶ See Binmore and Shaked, “Experimental Economics: Where Next?” and Fehr and Schmidt’s reply, “On Inequity Aversion.”

⁷ See Bicchieri, *The Grammar of Society*, chap. 3.

⁸ López-Pérez, “Aversion to Norm-Breaking,” p. 263.

⁹ See Bicchieri and Chavez, “Norm Manipulation, Norm Evasion.”

¹⁰ A reactive emotions account is developed in my “Priority of Social Morality” and “In Defense of (Some) Vainglory,” where I consider studies of the Ultimatum Game and a relative of it, the “Power-to-Take” Game.

in an egalitarian way—indeed, some small-scale societies play them very much as early economists predicted, with very low offers made by Proposers and accepted by Responders.¹¹ Strikingly, it is American university students who turn out to be some of the most egalitarian players in the world.¹² Once again, it seems that the truly outstanding feature of our evolved moral psychology is our ability to follow a wide variety of sharing and fairness norms in different circumstances and cultures.

Hayek, at times, suggests that perhaps this overestimates the continuity between small- and large-scale societies. For both Popper and Hayek, a fundamental contrast is between the “concrete” nature of the closed society and the “abstract” nature of the Open Society.¹³ This contrast has a number of dimensions: concrete societies are said to be face-to-face (as LPA societies surely were), while abstract societies interact with numerous unknown others according to abstract categories, whom we do not know “concretely.” For present purposes, the important distinction is that—at least at times—Hayek seems to argue that in a closed society rules serve known and concrete purposes: people follow rules because these rules serve known and specific ends, whereas in the Open Society rules are abstract. The Open Society’s rules are “purpose independent” insofar as they are followed without knowing what they do for us.¹⁴ If this were so, there would indeed be a normative gulf between LPA societies’ regulation by “concrete norms” and the Open Society subject to “abstract rules of order.” Now I certainly do not wish to deny Hayek’s fundamental point that in simpler orders we have a better grasp of consequences of our actions (§26), but we have seen that because social forms evolve through copying, even in simple societies people often do not possess correct causal models of the purposes their rules are serving, or how they do so (§11.1). Recall Joseph and Natalie Henrich’s analysis of food norms applying to pregnant and nursing women on Yasawa Island. The strict norms against eating large fish tracked the level of potential threats of ciguatera poisoning. Strikingly, however, contrary to the “known purposes” hypothesis, although women display consensus on what the rules are, when asked to explain the causal model the responses widely varied: women had learned the rule, but its “concrete purposes” were obscure. All cultural existence is about learning and internalizing a variety of rules without being clear what their “concrete purposes” are.

13.3. A Cultural Species

There is much more continuity between small-society social rule learning and the abstract rules of the Open Society than, at times, Hayek intimated. We again come to the conclusion that a critical part of the Modern Egalitarian Moral Package is

¹¹ For a now-classic study, see Henrich and Smith, “Comparative Experimental Evidence from Machiguenga, Mapuche, Huinca, and American Populations.” More generally, see the essays in Ensminger and Henrich, eds., *Experimenting with Social Norms*.

¹² Henrich, Heine, and Norenzayan, “The Weirdest People in the World?” and Henrich and Smith, “Comparative Experimental Evidence from Machiguenga, Mapuche, Huinca, and American Populations.”

¹³ Hayek, *The Political Order of a Free People*, p. 162; Popper, *The Open Society and Its Enemies*, vol. I, pp. 173–6. See also North, *Understanding the Process of Economic Change*, pp. 18, 70ff, 100.

¹⁴ Hayek, *Rules and Order*, pp. 50, 80.

“purpose-independent” normative guidance: it is this that allows culture to change swiftly as most of us exploit the benefits of the successful actions of others. As I have just stressed, a distinctive feature of humans is our ability to learn without understanding the relevant causal models. Interestingly, this reinforces another, far more important, theme of Hayek’s: the demand that one must have an accurate causal model before endorsing a rule is a relatively recent development, driven by an extreme valorization of a type of scientific, causal reasoning. A feature of what Hayek called “Cartesian Rationalism” is an insistence that only claims verified by the explicit use of reason are to be trusted.¹⁵ Many have criticized Hayek’s project on the abuse of reason in “scientism” in social science,¹⁶ but his fundamental point is vindicated by our understanding of humans as cultural creatures, who have developed an ultra-social life through our ability to reconcile on social and moral rules whose functions are often causally opaque to us. This goes deeper than the important claim that our system of rules was not designed (i.e., they are the result of human action, not design):¹⁷ even once the system has arisen, we very often cannot understand what purposes it serves, and what its effects are. This is not Hayekian obscurantism, but it is critical to the constitution of the life of a cultural species. In the next Part we shall see that, far from this causal opacity being overcome by advances in the social sciences, the explosion of social complexity has made it more intractable than ever.

That we are a cultural species is, then, yet another reason why the space of human social existence is so large. “Far from being trapped in the past by an outdated biological legacy, humans are characterized by a remarkable plasticity. Our adaptiveness is reinforced by both cultural and biological evolution.”¹⁸ Again, while we are by no means blank slates on which any culture can be written, our diverse nature and highly developed ability to learn new successful behaviors allow human society to change many of its features with amazing swiftness. As we have seen, humans have gone from deeply egalitarian social lives to remarkably hierarchical ones, and then back toward egalitarian societies—and all that from the despotic *Pan*.

Morality, Hayek argued, is a critical part of culture.¹⁹ As we have seen in this Part, there is overwhelming evidence for this claim. Yet some will press, what does this really mean? “[T]alk of morality is itself ambiguous. Do we mean morality as that which is done and enforced within a particular group? Or do we mean that which in some absolute sense simply ought to be done, regardless of group norms and loyalties? And what, if anything, is the connection between the two?”²⁰ There is certainly a distinction to be made between *de facto* and *de jure* morality.²¹ According to Kurt Baier,

we could not properly speak of a morality, as opposed to a system of conventions, customs, or laws, until the question of the correctness or incorrectness, truth or falsity, of the rules prevalent in a community is asked, *until, in other words, the prevalent*

¹⁵ Hayek, “Individualism True and False,” p. 53.

¹⁶ See his *Studies in the Abuse and Decline of Reason*, which includes *ibid*.

¹⁷ Hayek, “The Result of Human Action but not Design.”

¹⁸ Leland, *Darwin’s Unfinished Symphony*, p. 232.

¹⁹ See, e.g., Hayek, *The Fatal Conceit*, chaps. 1 and 2.

²⁰ O’Hear, *Beyond Evolution*, p. 101.

²¹ See Gaus, *The Order of Public Reason*, pp. 163ff.

rules are subject to certain tests. It is only when the current rules are no longer regarded as sacrosanct, as incapable of alteration or improvement, only when the current rules are contrasted with other, possible, improved, ideal rules, that a group can be said to have a morality as opposed to a mere set of taboos.²²

This is what I have called the “testing conception” of *de jure* morality.²³ At one juncture, Rawls claimed that in thinking about justice we do not seek to design practices from scratch—people already have existing rule-based practices—but we wish to know when legitimate complaints can be advanced against them.²⁴

This distinction between a society’s *de facto* morality and its justified *de jure* morality manifestly cannot be ignored.²⁵ A primary aim of the next Part will be to understand what type of justification makes sense once we appreciate morality as an evolved human cultural form. However, that a *de jure* morality must be distinguished from a *de facto* code does not imply that a *de jure* morality—a morality that passes the relevant test—is not also necessarily a part of functioning culture. The overriding claim of Part I has been that a shared morality has a fundamental role to play in human ultra-sociality: indeed, as I have stressed, it may well have been *the* critical development that allowed the transformation from *Pan* to ultra-social *Homo sapiens*. Regardless of how aesthetically pleasing a moral ideal may be to the philosophical mind, if it cannot adequately perform the function of coordinating our empirical and normative expectations, it cannot qualify as a *de jure* morality.²⁶

We might distinguish moral theory from moral philosophy.²⁷ Moral philosophy, we might say, inquires into the status of normative propositions, the nature of moral knowledge, and the construction of moral ideals and principles. I have nothing to say here about this field of inquiry: it is a form of regimented study with its own methods and criteria for success. In contrast, my concern is moral theory, focusing on morality as a social phenomenon, as having a function in human societies: it understands morality as the distinctive way that human social life is structured, and focuses on what is needed for it to perform this role in a manner that passes a Baier-like test to attain the status as a *de jure* morality. A philosopher can show that the *de facto* moral code of a society is not *de jure* (i.e., people have legitimate complaints against it), but she cannot create a *de jure* morality as a result of her philosophical investigations, just as she cannot create a *de jure* political regime simply by identifying a certain type of regime as the ideal. A regime that is just a dream or a proposal does not have *de jure* authority, because it has no authority at all. So too with a *de jure* morality. Moral theory is concerned with the justification of a social moralities that can “play their expected role in human life.”²⁸ As Rawls argued in his earlier writings, this may call for reform

²² Baier, *The Moral Point of View*, p. 174. To readers of my *Tyranny of the Ideal*, it should not be surprising that I set aside the “ideal rules” criterion. Emphasis added.

²³ Gaus, *The Order of Public Reason*, pp. 176ff, 424ff.

²⁴ Rawls, *Justice as Fairness*, pp. 52–4.

²⁵ Hayek, as I pointed out in §3, also at times seems skeptical of the distinction. *Rules and Order*, pp. 81ff; *The Fatal Conceit*, pp. 66ff.

²⁶ Of course, it is by no means restricted to this role. Automobiles have the role in human life of providing transportation, but this is consistent with stretch limos, Maseratis, Smart cars, and ’57 Chevy Bel Airs.

²⁷ See my “Moral Constitutions.”

²⁸ Rawls, “The Independence of Moral Theory,” p. 286.

of the existing practice, authorizing complaints against it, but it is a process of complaint and reform, not of ideal design.

13.4. The Push-Pull of Human Sociability

As we saw at the outset (§1), it is currently in vogue to claim that humans have “tribal” sentiments that pull us away from the Open Society. This is objectionable in a number of ways. For one, as soon as cooperation proceeds past close kin, there is a push outward–pull inward tension between expanding the sphere of cooperation and nepotism. The rise of clans can be seen as a push outward to wider cooperation, extending the idea of kin; yet it was also a system that simultaneously pulled members back toward these extended kin relations and away from a more impartial morality. The much abused “tribes” were often a way of expanding clan-based networks, but like most expansions, they also drew boundaries—wider boundaries, but boundaries nevertheless. The development of tribes, like kingdoms and later nations, pushed us outward to wider networks of cooperation while pulling us inward to richer social and moral relations within the group, while kin relations continued to exercise their strong centripetal force. Extending core morality beyond kin-based networks may have been the critical development of WEIRD morality, but WEIRD morality too manifests a push toward expansion of the impartial network and pulling back by kin and ethnic markers and the power of social proximity. Human social life is defined by this constant tension between the push to wider moral relations and the pulling back of familiarity and social proximity. To describe human morality as either tribalistic or an ever-expanding circle is evocative but fundamentally distorting.

A greater distortion implicit in the “tribal” claim is the supposition that there is a “tribal instinct” that explains our inward-pulling inclinations. I have endeavored to show here that a number of capacities and inclinations make up the “Tribalistic Package,” and so we can expect the inward pull of social proximity to be strongest when they align. Consider the various dimensions of “tribalism”:

- Ethnic markers are part of learning norms and cultural learning by imitation.
- Trust seems related to social proximity, and the knowledge that others share one’s normative network. Trust is, in turn, critical to collective action to promote public goods.
- Norms tend to erode in the face of knowledge of others’ noncompliance; social proximity tends to retard this erosion, and so strengthens norms.²⁹
- In times of conflict, pro-social behavior is intensified, though this too may well be mediated by more demanding pro-social norms.
- Humans seem able to adopt a stance of collective intentionality—a “we,” as opposed to an “I,” perspective on action.
- Human morality itself requires that we “take sides” in disputes. As we saw in our strategic model, this can lead either to “tribal” signaling that one is a member in good standing of “our group” or to an impartially expanding morality.

²⁹ Bicchieri, Dimant, Gächter, and Nosenzo, “Social Proximity and the Evolution of Norm Compliance.”

- We must remember that *Pan* was a coalition-forming, despotic, species. In humans this can lead to coalitions in which some seek to dominate others.
- We are constantly drawn to pay special attention to kin and kin-based parochial organization.

When these line up in a way that strongly pulls us inward, we are very “tribal” indeed. Communitarian-egalitarians who wish to show our collective inclinations understandably look to the Second World War—an apparent struggle to the death between moralistically opposed ethnic groups. This was not a resurgence of “the tribalistic instinct,” but something like a perfect storm of inward-pulling forces. Yet, in a characteristically human way, this almost immediately gave way to the more complex interplay of the expansion of cosmopolitan norms and moralistic conflict of the postwar and postcolonial era.

Are we fit for the Open Society? Of course we are. We are certainly not inherently groupish creatures who readily turn their backs on cooperative expansion. Human history since the Late Ice Age has been the gradual drawing out of humans through a variety of institutional arrangements—including morality—at each step being pushed out and pulled back. The distinctive feature of the Open Society (and its basically WEIRD morality) has been to give this expansion a more liberal-egalitarian trajectory. Yet Hayek’s claim that our “egalitarian sentiments” clash with the Open Society reminds us of an important truth: we are not optimized for it—as we have not been optimized for any social order, even, I would venture, Late Pleistocene Egalitarianism. Ants are optimized for ultra-social living, not the descendants of despotic primates.

PART II

DIVERSITY AND SELF-ORGANIZED COMPLEXITY

Is the Open Society beyond Justification?

This Part takes up Hayek's second unsettling thesis: Is the Open Society too complex for the practice of moral justification? To begin to analyze this unsettling thesis, we must first consider the nature of the Open Society, and why it is characterized by extreme diversity and complexity. This inquiry thus begins by distinguishing Millian and other liberalisms from the far more deeply diverse Open Society. I shall argue that the Open Society is characterized by a process of autocatalytic diversity, which leads to ever-increasing complexity. Having analyzed the Open Society, I begin to focus on Hayek's second unsettling thesis by first examining the nature of order in this highly diverse Open Society: here I shall draw extensively on the results of the first inquiry, arguing that the self-organization and morality of the Open Society are based on the Modern Egalitarian Package. This also completes the examination of the first unsettling thesis, arguing not only that we are fit for the Open society, but that the order of the Open Society depends on the evolved Modern Egalitarian Package. This second Part concludes with an account of how the constitutive moral rules of the Open Society can be justified. I endorse Hayek's criticism of social contract theory, proposing in its stead a self-organization model of moral justification.

§14 Liberalism and the Open Society

14.1. What Is the Open Society?

Part I addressed Hayek's thesis that our evolved moral nature is in fundamental conflict with the cultural evolution that has produced the Open Society. These evolutionary investigations not only allowed us to evaluate Hayek's worry, but also, we shall see, provide us with resources to think about the Open Society and justification within it. We now turn to Hayek's second unsettling thesis: that the Open Society is too complex for us to justify its constituent rules. The foundation for this second thesis has been laid in our analysis of the causal opacity of cultural rules (§11.1): if we cannot grasp what our norms accomplish, it is not clear what would count as justifying them. Still, some may think that advances in social science and better causal models in scientifically inclined Open Societies will allow us to see through this opacity and uncover the effects of our rules. Perhaps, in a Deweyan spirit, we might be able to apply our scientifically informed intelligence to better understand the Open Society. Hayek's thesis is just the opposite: however serious a problem causal opacity is for cultural rules generally, he sees it as exponentially greater in the Open Society.

To explore this second Hayekian unsettling thesis, we must first understand what "the Open Society" is. It is tempting to see it as simply another name for a society with extensive liberty to discover and innovate—Millian liberalism in Popperian guise. Mill's *On Liberty*, one might suppose, is the catechism for the Open Society. This Part is an extended argument against this supposition, arguing that a "Great" or Open Society is a distinctive form of liberalism—and one that holds out the best hope of a free and diverse society. Yet we shall see that Hayek's core point is sound: given the nature of the Open Society, normative justification is a far more daunting task than political philosophy has realized. However, I hope to show that our justificatory resources are richer than Hayek was apt to think.

The first step, then, is to distinguish the idea of an Open Society from a progressive liberal order such as Mill had in mind. Having done that, in the remainder of this Part, I move to further investigating the deep complexities of the Open Society, before finally exploring the philosophical problem of the nature of justification within such an order. Again, my aim is to show Hayek's prescience in grappling with problems of which his contemporaries had not the faintest glimmers, while hopefully providing a more robust justification of such an order.

14.2. Millian Liberalism

One of Hayek's weaknesses was, alas, his tendency to Manichean dualisms in his analyses of the history of moral and social thought, such as that between "Cartesian rationalism" and the thinkers of the Scottish Enlightenment. Besides their Manichean tone, it usually ended up that the French were on the side of darkness. (I have already,

no doubt, enraged some loyalists by invoking Rousseau as a model in Part I.) Yet for all that, Hayek was undoubtedly correct about a basic point: he is following a tradition of thinking that rejects what, on Michael Freeden's excellent analysis, are core commitments of mainstream liberal thinking. "It is," Freeden observes, "virtually a truism to state that liberals view human nature as rational."¹ Freeden uses Mill as an exemplar, reminding us that Mill approvingly quotes von Humboldt (who also provided *Liberty's* epigraph), "the end of man, or that which is prescribed by the eternal or immutable dictates of reason, and not suggested by vague and transient desires, is the highest and most harmonious development of his powers to a complete and consistent whole."² Hayek, to be sure, overuses the term "rationalism," and certainly on many dimensions Mill was not a rationalist, but Freeden is correct that when it comes to Mill's account of human and social development, the powers of individual reasoning are fundamental. In this he is akin to Dewey, who also depicted intelligence and problem-solving abilities as core attributes of individuality and social progress.³ Thus for Mill the growth of the "powers and acquirements of mind" is a key factor in the rise of "civilization."⁴

This brings us to another of Freeden's core features of the liberal project: a conception of humans as "progressive beings."⁵ As humans increasingly develop their powers of reasoning, civilization and morality advance. Reason interrogates existing customs and practices founded on force and superstition: when it cannot find a justification, they are to be rejected in favor of those based on freedom and equality.⁶ Social and moral theory is, to a significant extent, an ongoing exposé of what is arbitrary and irrational, producing an ever-more rational morality and society among ever-more rational persons. The progressive commitment of this Millian-Deweyan rationalism is perhaps best manifested by Kitcher's ethical project: rational inquiry has generated not simply reform and change, but a progressively better morality.⁷

Following Freeden's analysis, we thus can identify two (there are other) features of this "Millite core" of liberalism: (i) the valorization of reason and rationality, both as a basic element of the individual good and as a tool for social criticism and reform; and (ii) a belief that the application of rationality and inquiry leads to social and moral progress. While these commitments manifest the importance of rationality to this liberalism, both Mill and Dewey stressed that it is not a dogmatic rationalism but a fallibilistic and tentative reasoned inquiry. It recognizes that moral and social philosophy rely on inconclusive data and address complicated problems. The task of ethical inquiry is the application of critical reason to the evaluation and reform of existing social institutions and practices. Free reasoning individuals criticize superstition, taboos, arbitrary customs and norms, seeking to replace them with rational practices that better suit human needs and manifest freedom and equality. While always looking at existing arrangements critically, it is also cautious about the powers

¹ Freeden, *Ideologies and Political Theories*, p. 148.

² Mill, *Liberty*, p. 261; Freeden, *Ideologies and Political Theories*, p. 148.

³ See further my *Modern Liberal Theory of Man*, p. 28 and accompanying notes.

⁴ Mill, "Civilisation," p. 121.

⁵ Mill, *On Liberty*, p. 224.

⁶ This is most manifest, I think, in Mill's *The Subjection of Women*.

⁷ Kitcher, *The Ethical Project*, pp. 138–249, 285–330.

of reasoned inquiry. This caution leads it to reject utopian schemes for improvement in favor of piecemeal progress arrived at through free inquiry, always revisable in the light of new evidence.⁸ This is of first importance in giving the doctrine its liberal, “open,” feature, stressing free inquiry and tolerance of different views.

14.3. Diversity, Discovery, and Progress

Understandably, then, some of the best political philosophers who take human diversity seriously embrace aspects of Millian liberalism. Its openness to diversity and rational inquiry is wonderfully manifested in Mill’s idea of “experiments in living.”⁹ As Ryan Muldoon sees it, “Mill’s approach to experiments in living offers us an account of *social discovery*. On this kind of account, justification becomes subsumed to iterated discovery, which includes a permanent competition of perspectives.”¹⁰ Reaffirming Popper’s criticism of utopian thought, Muldoon argues that on the Millan approach, instead of identifying “a regulative ideal and comparing ourselves against that a priori standard, we try competing approaches out, *and see what works in our circumstances*.”¹¹ Thus discovery and experimentation have “primacy” for political justification in a diverse changing world.¹²

This language of “experiments” and “discovery” points to a scientific analogy, which Muldoon carefully develops, drawing on current models of the division of cognitive labor in scientific inquiry.¹³ These models assume diverse teams (or individuals) exploring different parts of an agreed-upon “scientific landscape.”

Scientists are imagined to be “hill-climbers” on an unknown “landscape.” The landscape itself is interpreted as a topic of scientific inquiry. The X and Y dimensions represent potential research approaches. *The Z dimension represents the epistemic significance of any findings to be had given the research approach indicated by the (X, Y) position.* . . . At the beginning of inquiry, scientists have no knowledge of the landscape—that is, they do not know anything about the comparative significance of any research approaches. *They discover this only by traversing the landscape.*¹⁴

These models assume that when one competent scientist reports the Z value of a specific coordinate (X_i, Y_j), others generally concur, and so everyone can benefit from the searches of others. They are exploring essentially the same landscape in different ways: in this sense, there is a general agreement on what would constitute progress,

⁸ Popper, *The Open Society and Its Enemies*, chap. 9.

⁹ See Müller, *Polycentric Democracy*. See also Kitcher, *The Ethical Project*, chap. 3; Anderson, “John Stuart Mill and Experiments in Living.”

¹⁰ Muldoon, *Social Contract Theory for a Diverse World*, p. 30. Muldoon presents a multistage account of social morality; at later stages, his view shares much with that developed here.

¹¹ *Ibid.*, emphasis added.

¹² *Ibid.*, p. 35.

¹³ *Ibid.*, p. 30. Kitcher, another proponent of experiments in living, was one of the first to explore the cognitive division of labor in science. See his “The Division of Cognitive Labor.”

¹⁴ Muldoon, “Diversity and the Division of Cognitive Labor,” p. 120. Emphasis added. For a more thorough exploration of this idea, see §26.1.

i.e., a better solution is one higher on the Z axis. Scientists who are members of the same scientific community are exploring the same epistemic landscapes—they share the similar problems, tools, standards, and categories such that when one discovers a solution to problem *P*, others will generally agree that it is indeed a solution to *P*.¹⁵

The distinctive feature of this Millian liberalism-as-inquiry is a defense of openness and diversity *as a tool to generate better answers to our moral and social problems*. Progress is measured in terms of climbing the moral and social landscape—getting better solutions to problems of social and moral life. Society is conceived as an array of diverse experiments in living that are exploring a common landscape to discover better ways of living together. Under these conditions, we should expect that a diverse group of explorers, engaging in different and competing experiments in living, will more efficiently and thoroughly explore this social landscape, leading to better and better solutions to our problems.

Although Millian liberalism is a strong defense of diversity and disagreement, it also implies somewhat severe limits on the range of useful diversity. There cannot be progress if explorers continue to investigate parts of the landscape that have been shown to be barren of new insights. The essence of science is that the community gives up on some regions and comes to focus on the more promising ones: those who refuse to do this lose their good standing as members of the scientific community. Those who refuse to accept the findings of experimental results showing the way to moral progress must demonstrate a sort of moral incompetency—their view has failed the experimental test.¹⁶ This is not to say that these “incompetents” will be prevented from pursuing useless experiments. Like Mill, we may welcome palpable error as a way to remind ourselves why it is so erroneous.¹⁷ But a claim to liberty to explore these discredited ways of living cannot be based on their manifestation of core values on individuality, reason, and progress. The point of experimentation is to learn, and to learn is to discard discredited claims. It is simply incoherent to adopt the progress-through-experimentation approach and yet argue that all options remain as viable and worthy alternatives. Citizens who continue to reject the discovered truths must in some way be judged “unreasonable,” and their views discounted in public justifications.¹⁸ If there is progress, some disagreements simply must be put to rest: ethical and social theory must conclude that some ways of living are not of value and cannot be taken as serious options in a progressive morality. To that extent, the social-moral debate is over.

Millian defenders of the Open Society may invoke a “Kuhnian” option to escape this constraining result.¹⁹ In periods of normal science, the community agrees on the scientific landscape and how to explore it, but in revolutionary periods, the scientific community fractures, with different understandings of the epistemic

¹⁵ We can understand a scientific paradigm as articulating an epistemic landscape and the tools to explore it. See Kuhn, “Second Thoughts on Paradigms.”

¹⁶ Anderson, “John Stuart Mill and Experiments in Living.”

¹⁷ *On Liberty*, chap. 2.

¹⁸ This is the view of Quong, who, though he self-identifies with Rawlsian liberalism against so-called perfectionism, shares many of the latter’s commitments. See his *Liberalism without Perfection*, chap. 10. Cf. my “Sectarianism Without Perfection?”

¹⁹ Kuhn, *The Structure of Scientific Revolutions*, pp. 111–74.

landscape—what the problems are and what constitutes superior solutions. So different communities are exploring different landscapes. As I have argued elsewhere, the Open Society is characterized by many such communities of moral inquiry, each exploring a different landscape.²⁰ In this sense, the Open Society might be thought of as a society in a perpetual and radical state of “moral revolution,” with different paradigms constantly competing. However, in this case the analogy with science breaks down, for the problem-solving, progressive nature of science requires an end to revolutions to get on with the shared problem-solving that defines progress. Without that, there are simply many competing schools, each disputing each other’s “data” and theories—as is the case in moral and political philosophy. I am not claiming that this is an unfortunate condition, but it does not allow a political framework such as Millian liberalism to claim that it spurs moral progress in a way even roughly similar to science: under conditions of deep diversity, there is no shared community of moral inquiry on which to base this result.

14.4. The Progressivist Attack on Superstition: False Beliefs

Understood as social experimentation aiming to discover the *better morality for all*, the great promise of Millian-Popperian-Deweyan liberalism is that openness, diversity, disagreement, and liberty are integrated into a theory of moral progress according to which we truly discover better liberal arrangements. Nevertheless, as Hayek argued, this program has a rationalistic slant that renders it limited as a framework for a diverse society, for progress inherently restricts the range of sensible diversity. A critical example is religious belief. Almost all Millian liberals defend freedom of religion, but not because the religious are pursuing valuable experiments in living. An enduring theme in Millian-Deweyan progressivism is the irrationality of religion, and the threat it poses to rational inquiry. Thus in his 1939 discussion of the place of science in a free culture, Dewey claimed:

The historic influence of religion has often been to magnify doctrines that are not subject to critical inquiry and test. The cumulative effect in producing habits of mind at odds with attitudes required for maintenance of democracy is probably much greater than is usually realized. Shrewd observers have said that one factor in the relatively easy victory of totalitarianism in Germany was the void left by the decay of theological beliefs. Those who had lost one external authority upon which they had depended were ready to turn to another that was closer and more tangible.²¹

Dewey is, perhaps, blunter than most, but some such view is implicit in the scientifically inclined critical basis of Millian liberalism. Those who ignore science and rational inquiry to build their lives on discredited ideas are like alchemists: they may be free to proceed with their studies, but they merit derision in a progressive society.

²⁰ See my *Tyranny of the Ideal*, chap. 3.

²¹ Dewey, *Freedom and Culture*, p. 151.

And as Dewey is forthright enough to point out, they are apt to be a threat, because they have turned their backs on the very spirit of critical inquiry on which the Millian liberal order depends.²² In a fundamental sense, one simply cannot be both religious and a good liberal.

Millian liberalism is thus an inherently a secular theory: the reasoned and critical stance it so valorizes, and on which moral progress itself is based, requires a public social world of facts and values, just as science requires a common world of facts and the shared values of inquiry. In this secular world, religion is ultimately out of place, and needs to be reinterpreted to intelligibly fit. In Amartya Sen's version, religion becomes a case of participation in an ancestral culture, something that at least makes sense in a liberal world of secular facts.²³ Alternatively, religion may be interpreted in a solely social functional sense as providing "community" or "belonging." But what is so elusive to the Millian, progressivist liberal is understanding religion on its own terms. Hence it struggles to make sense of an extraordinary amount of real diversity—after all, most people in the world *are* religious.²⁴ Again, this is not an error in articulating the view: there cannot be a commitment to moral progress while beliefs that our experiments have shown to be false continue to be treated on par with those backed by evidence or, at least, those which seemingly can in some way be rationally vindicated.²⁵

The problem becomes sharpest when religious belief runs contrary to progressive principles of justice. Consider gender equality, almost always considered an aspect of our permanent moral progress.²⁶ Religions, alas, often affirm gender inequality, especially as regards the family, which must be seen by Millian liberalism as unjust. Understandably, liberals are torn. One possibility is, on the grounds of religious freedom, to allow unjust family relations based on religious convictions, whereby children are raised in patriarchal families. But this implies the uncomfortable position of protecting a right to do what is wrong.²⁷ "The family, justly constituted, would be the real school of the virtues of freedom."²⁸ There must be something suspicious about people claiming a right to do what moral progress has shown to be wrong—and on the basis of false beliefs. It makes the Millian liberal doubt whether religion should be accorded insulation from public justice—the progressivist liberal may simply deny any such insulation, allowing the liberal state to intervene in family life and religious affairs to promote public justice.

²² For broadly similar worries about Deweyan democracy, see Talisse, *A Pragmatist Theory of Democracy*, pp. 44–53. Talisse looks to other elements of the pragmatist tradition to resolve this problem.

²³ Sen, *The Idea of Justice*, p. 237.

²⁴ According to the Pew Center, "By 2055 to 2060, just 9% of all babies will be born to religiously unaffiliated women, while more than seven-in-ten will be born to either Muslims (36%) or Christians (35%)." <https://www.pewforum.org/2017/04/05/the-changing-global-religious-landscape/> (accessed July 2019).

²⁵ In a university meeting on the value of diversity, I suggested including religious diversity. The response was, roughly, "but they're nuts!" And of course my fellows were correct in their resistance: on Millian progressivist liberalism, we might as well include alchemists.

²⁶ See Kitcher, *The Ethical Project*, pp. 145–53; Buchanan and Powell, *The Evolution of Moral Progress*, pp. 6, 12.

²⁷ See Waldron, "A Right to Do Wrong."

²⁸ Mill, *The Subjection of Women*, p. 295.

14.5. The Millian Attack on Superstition: Unsupported Beliefs

According to Stevie Wonder, “When you believe in things that you don’t understand then you suffer superstition.” Millian liberalism’s inherent suspicion of custom reacts to its unreasoning nature. In the early paragraphs of *The Subjection of Women*, Mill notes how difficult it is to convince the late-nineteenth-century public of the injustice of its norms. Mill insisted that his reasoned argument must confront “barbarisms” of tradition and custom that are devoid of rational basis, depending only on instinct and feeling.²⁹ I would imagine few if any readers today would have any qualms with Mill’s position. Like Mill, almost all liberals today are suspicious of custom as the accretion of past injustices, inequalities of power, or, as Mill says, barbarism.³⁰ In this sense, however, Millian progressive liberalism is fundamentally hostile to culture itself,³¹ for we have seen (§11.1) that the essence of culture is learning causally opaque social rules through imitation. Of course each person generally will have individual aims, desires, and beliefs when acting on cultural rules, but as we have seen in the case of the Yasawa women, these seldom capture what the rule is really doing for us (§11.1). To reiterate: critical to a cultural species is the ability to learn by copying what one does not really understand—and that’s superstition. It is precisely what rational progress is supposed to undermine. Progress involves an increased use of reason to replace reliance on unreflective custom. While the Yasawa adopt rules with unknown purposes, progressive (Western?) moderns do not. Although cultural evolution does not render us unfit for the Open Society, we may well be unfit for progressivist liberalism. As we saw in Part I, humans have achieved ultra-sociality through culture. When an innovation proves successful, others need not understand the reasoning behind it and why it is functional. Indeed, the innovator herself need not, and most likely did not. If all sought to take up the Millian stance, culture would collapse: we act on cultural rules because we have learned by observation of our fellows. To interrogate all rules—to insist that what cannot be understood is to be distrusted as superstition or prejudice—is directly opposed to our nature as a cultural species. Our truly distinctive trait, recall, is not individual reasoning but social learning (§§11.1–2). The progressivist program is a recipe for drastically reducing social learning (aka imitation), throwing us back on our cognitive capacities.³² That, however, is in turn a recipe for undermining

²⁹ Ibid., pp. 261–3.

³⁰ As I note in Part III (§28.4), some progressives continue to use this term to describe the practices of traditional societies. Recall its dictionary definition: the practice of barbarians, i.e., those who are “savage, heathen, brutes, or beasts.”

³¹ Being Mill, he tries to find a role for custom and tradition in a progressivist liberalism: “The traditions and customs of other people are, to a certain extent, evidence of what their experience has taught *them*; presumptive evidence, and as such, have a claim to his deference. . . . Customs are made for customary circumstances, and customary characters; and his circumstances or his character may be uncustomary. . . . *He who does anything because it is the custom, makes no choice. He gains no practice either in discerning or in desiring what is best.* The mental and moral, like the muscular powers, are improved only by being used.” *On Liberty*, p. 262 (emphasis added). Again note the inconsistency with culture.

³² As indicated in the quotation in the previous note, Mill sometimes suggests that “customary characters” follow custom, not the uncustomary ones. An elitism or defense of a “clerisy” is never far from the surface of progressivist reason: after all, it is the expert scientific minds who are best trained at critical inquiry. While innovators are needed in any culture, the division between the free-thinking elite and the customary masses is a familiar theme in progressivist liberalism. On the clerisy and its rejection of the bourgeois commonplace, see McCloskey, *Bourgeois Equality*, chaps. 62–3.

ultra-social cooperation, and would probably make any significant system of social rules dysfunctional. Indeed, I shall presently argue that progress itself depends on our ability to copy.

Alas, unfortunately many readers will take away from this discussion that I believe public norms supporting gender inequality can be justified, or are legitimate, in the moral constitution of the Open Society. I think nothing of the kind, as we shall see (§20). The subtler point is that a rejection of customary norms and rules simply because they are ungrounded in reasoned analyses, inquiry, or argument is so sweeping a critique that most of culture would then be rejected, as each is left to her own (not overly impressive) cognitive devices. The only viable way for the Millian progressivist to escape this charge is to maintain that, with the advent of a rational, scientific culture, what was previously shrouded in obscurity now can be made transparent. Perhaps the “tribal” “closed society” is characterized by “taboos” that are rationally opaque, but not the open, critically rational, society.³³

The remainder of this Part is a defense of Hayek’s claim that, so far from this being the case, the functioning and consequences of our social rules are increasingly opaque. The Open Society is increasingly diverse and complex. Indeed, I shall argue that diversity and complexity are *autocatalytic*: increased complexity leads to increased diversity, which generates more complexity and more diversity, leading to runaway complexity. Given this, the program of Millian-Deweyan progressivism is increasingly implausible: it certainly cannot be applied to the social world we have created—and this may have been the case for thousands of years.

14.6. The Paradox of Justification in the Open Society

As Hayek believed, a social theory that puts rational criticism and understanding at the heart of the moral evaluation of social rules fails to appreciate the culturally based social life of humans. It is hard to overestimate the importance of his insight. This is not, as some readers may be apt to conclude, to attack rationality. Defenders of rationality and rational criticism are not those who adamantly insist that all social life is within their scope and thus requires of them more than they can possibly deliver. Rather, they appreciate both their powers and limits and, within those limits, show how they can be most effectively employed. This marks the distinction between a liberal and a traditional conservative.³⁴ Conservatives tend to understand cultural life much better than liberals, appreciating its path-dependent and opaque mechanisms. In the face of the opacity of much culture, some conservatives recommend an attitude of respectful acceptance, with change only at the margins.³⁵ This, though, can lead the conservative to endorse a public morality that consistently and systematically treats some as inferiors and others as superiors—to the liberal, a quintessentially oppressive

³³ See Popper, *The Open Society*, vol. I, pp. 15, 60.

³⁴ This is to be distinguished from the “fusion conservatism” that has characterized American politics for the last half century. See my essay, “The Role of Conservatism in Securing and Maintaining Just Moral Constitutions.”

³⁵ “As has often been acknowledged by conservative writers, one of the fundamental traits of the conservative attitude is a fear of change, a timid distrust of the new as such....” Hayek, *The Constitution of Liberty*, p. 400. Cf. Oakeshott, “Rationalism in Politics.”

relation. The liberal better understands the dynamics of morality, and how its functioning is grounded on widespread sensitivity to moral rules, and so public justification (§9.3).

Yet the function of moral rules is typically opaque to us. We thus commence our analysis facing a paradox: the liberal seeks public justification of our rules, the consequences of which are uncertain, the function of which are unclear. We have been exploring one response: the liberal may follow Mill and Dewey in subjecting all social forms to the critical interrogation of reason. If forced to choose between a commitment, on the one hand, to reason and progress and, on the other, to a truly diverse and complex Open Society, many, perhaps most, liberals are ready to opt for the former and abandon the latter. If a commitment to reason and progress is inconsistent with the complexity of our culture, and so the degree of diversity that is inevitably characteristic of the Open Society, then so be it: we must limit justification to “reasonable” diversity. As a student once remarked to me when confronted with the observation that many forms of diversity can be unpalatable and uncomfortable, “so much the worse for diversity.” Hence the temptation of the liberal to abandon the Open Society, and its messy and its often unpalatable diversity (think of an atheist passing a *parochial* school, or a devout Muslim walking down the Santa Monica boardwalk). In the spirit of our age, the Millian’s defense of diversity becomes an advocacy of a normalized diversity, dismissing many ways of living in the name of inclusiveness. In this Part, I follow another path—to understand both the sources of a much more radical diversity, and how this diversity can ground the impartial moral order of the Open Society. As my student recognized, such a society may be uncomfortable and may challenge moral certitudes. Yet, I shall argue, only the Open Society allows for a publicly justified morality in the midst of radical and ever-deepening diversity.

§15 Understanding Diversity

15.1. Modeling Agential Diversity

I have been arguing that Millian progressivist liberalism inherently “normalizes” diversity, requiring a distinction between reasonable disagreement about matters that remain open to inquiry, and the unreasonable who continue to propagate views that have been shown to be erroneous. I have thus far relied on an intuitive idea of diversity. In order to appreciate the depth and range of possible diversity, it will be helpful to sketch a more formal account.

Recent analyses of diversity have employed the idea of a perspective.¹ For present purposes, we can model agent *i* in judgment-action relevant contexts as possessing a perspective that has six elements.

¹ Muldoon, *Social Contract Theory for a Diverse World* and his “Perspectives, Norms and Agency”; Müller, *Polycentric Democracy*; Page, *The Difference*, chap. 1; Landemore, *Democratic Reason*, chap. 4. I have employed the idea extensively in *The Tyranny of the Ideal*, esp. chap. 2.

- I. A set of goals, evaluative standards, values, or normative principles.
- II. A method for modifying these standards, etc.
- III. A set of categorizations of the natural and social world. These categorizations determine the relevant features of the world (e.g., what are the social and natural entities, actions, states of affairs, etc.) for any judgment-action context, J_x . Categorizations thus determine judgment-action contexts. A description of an action ("φ is an action that involves harm to another") or a judgment ("φ harms others") is itself a categorization.
- IV. A set of options $\{\alpha, \beta, \gamma \dots\}$ in any given judgment-action context, J_x .
- V. A set of predictive models, which take the set of options $\{\alpha, \beta, \gamma, \dots\}$ and predicts for each, the social world(s) $\{S_1, S_2, S_3, \dots\}$ that would be produced (by each option). Social worlds are partially defined by the perspective's set of categorizations.
- VI. A function, which takes I, III, IV, and V and applies them to the resulting social worlds $\{S_1, S_2, S_3, \dots\}$, yielding a ranking of at least some worlds. A perspective may generate multiple non-aggregated rankings based on different evaluative standards (I), e.g., an interest-satisfaction ranking, a justice ranking, and so on.

This is, of course, a highly artificial model of agential judgment and choice—it is not intended as a psychological theory of how people actually choose. Many of the conditions could be relaxed. For example, the "function" in VI might be no more than "an intuition" or a guess.² The aim is to get clearer about the moving parts of an agential perspective, where the agent has a set of standards, aims, etc. (and some idea about how to modify them in the light of new information); looks at the world, identifying contexts for decisions; in any given context, examines the options, considers the social world that each option would produce; and makes a decision/judgment about the choice worthiness of the various options.³ The model allows us to dissect agential choice in a way that alerts us to the possible loci of diversity.

Diversity can arise from differences in any of these elements. Too often, diversity is thought to arise only from disagreements about element I, our basic aims, values, and principles. This has been the traditional focus of analyses of disagreement from economic decision-making to political morality.⁴ But the sources of diversity are much more diffuse. Once we appreciate the diversity of sources of perspectival diversity, we can see the futility of the idea that disagreement can be overcome if, somehow, we could all come to share the same set of values and ends—or if, a wee bit more realistically, disagreement about these matters could be greatly reduced in a true human community.⁵ Suppose that on each of the six elements there are ten reasonable

² The idea of mapping preferences to outcomes and so ranking outcomes is central to utility theory. See my *On Philosophy, Politics and Economics*, pp. 34ff.

³ Note that the concept of a "belief" does not enter in as a single variable, but is included via categorizations, predictive models, action options, and so on.

⁴ Many economic models implicitly suppose a homogeneity of preferences, but when disagreement is modeled, it is typically in terms of differences in preferences, though differences in beliefs and the mapping function can be critical. See my *On Philosophy, Politics and Economics*, pp. 30–5. On political pluralism, see Crowder, "Pluralism." For a nuanced discussion, see Landemore, *Democratic Reason*, pp. 213–9.

⁵ A common aspiration in communitarian literature is to transform society into a collective enterprise association, in which social order is structured around pursuit of common goals. For insightful analyses, see Oakeshott, *On Human Conduct*, Parts II and III; Benn, *A Theory of Freedom*, chap. 12. Although Rawls

positions, and that all combinations of the reasonable are themselves reasonable: that yields 1,000,000 reasonable perspectives! Now if we make the wildly strong assumption of *complete agreement* on evaluative standards (something not even characteristic of simple societies, §9.3), we still have 100,000 distinct perspectives. Omit element II, and we still have 10,000 perspectives.

15.2. Diversity of Categorizations

Those who resist this plethora of diversity are apt to deny that there are so many sources of respectable diversity. It would be tedious to canvass all the elements of a perspective to show their importance; let us focus on what may seem the most contentious: diverse categorizations.

One cannot, I would venture, even begin to understand the intractable dispute about the ethics of abortion unless it is firmly kept in mind that the protagonists categorize the world in deeply different ways. Many opponents of abortion rights understand an embryo to be ensouled: their world contains the category “being with a soul” that, on their moral perspective, is intimately linked to rights of a person. Proponents of abortion rights insist that an embryo is not a person, employing a naturalistic conception of a person that has no room for the category “ensouled” (or a number of other related categories). Hence the utter futility of this long, very angry, dispute, which still would be intractable if all the parties accepted the same set of evaluative standards. Or consider the categorization “homophobe,” employed by most “progressives.” Conservative criticism of homosexuality is categorized as a mental disorder (a phobia), indicating that the criticism is irrational and outside bounds of rational disagreement. On the other hand, the conservative often invokes the category of an “impure” or “blasphemous” action—categories that have little or no place in a progressive’s understanding of the world.⁶

There is strong evidence of cross-cultural differences in categorizations. Between the ages of seven and ten, urban children in the United States, for example, come to see humans as the exemplar animal from which sound inferences to other animals are made. “(1) Inferential projections of properties from humans are stronger than projections from other living kinds; (2) inferences from humans to mammals emerge as stronger than inferences from mammals to humans; and (3) children’s inferences violate their own similarity judgments by, for example, providing stronger inference from humans to bugs than from bugs to bees.”⁷ Urban Americans’ categorizations of animals are driven by shared perceived commonalities—a robin is similar to a sparrow because they share similar observed and structural features. Having very little knowledge of specific animals, for urbanites perceptual and structural

is careful to distinguish a political order from both a community and an enterprise association, his emphasis on civic friendship has led some of his followers to take a distinctly communitarian turn. See *Political Liberalism*, pp. xlix, 40–3; Van Schoelandt, *The Community of Public Reason*.

⁶ See Haidt, *The Righteous Mind*, chap. 6. When it comes to the environment and food consumption, notions of purity may, after all, be invoked by the progressive.

⁷ Henrich, Heine, and Norenzayan, “The Weirdest People in the World?” p. 67.

similarities (especially to humans) drive categorizations. In contrast, in small-scale and agricultural societies, in which there is a deep knowledge of fauna, animals may be categorized by their cultural importance: animals that are symbolically similar may be classified as similar animals despite varying greatly in their perceived features. In these contexts, inference patterns about features of animals is not anthropocentric—humans are not the key basis for inferences about other animals.

Differences have also been found in the categorization of spatial relations.⁸ Speakers of Dutch and English, for example, use the speaker as a reference point (“X is to the right of me”), while most languages categorize space in terms of directions that are in some way anchored to external objects (“X is to the right of object O,” “X is behind object O”). In one experiment this led to differences in judgments about the identity of objects. Dutch and Tzeltal speakers were first seated next to an arrow that pointed right (north), and then moved to another table sitting on the opposite side, and were asked which of two arrows was similar to the first they observed. The Dutch chose the arrow that was still to the right of them, the Tzeltal, that still facing north.

More generally, there is reason to think that even the categorization of individual agency varies culturally.⁹ Westerners are far more apt to categorize agency in terms of internal features of the agent, while others stress relational features. As Henrich, Heine, and Norenzayan conclude, there is considerable evidence for Clifford Geertz’s view that the very concept of the self or human is culturally laden: how we go about categorizing agents is itself part of a perspective.¹⁰

15.3. Moral Progress through Reducing “Shallow” Disagreement: The Illusion of Naïve Naturalism

On a Millian-Deweyian inspired view of moral progress, differences in categorizations are best understood as deriving from biases and errors. As experiments in living proceed, our better understanding of the social world and the elimination of biases will lead to increasing moral agreement—a crucial characteristic of moral progress.¹¹ On this view, humans do not radically disagree in evaluative standards, but are often confused by false theories of the social world and various cognitive biases. As we overcome these through experiments in living and the progress of science, moral diversity too is reduced.

Let us grant, for sake of analysis, the hypothesis that there is not a great deal of basic difference in evaluative standards—in the parlance of political philosophy, disagreement is not “deep” because it does not derive from basic moral disputes; rather, disagreement is caused by different ways of understanding the world. The question is whether such disagreements are any “shallower” or more tractable than so-called

⁸ Ibid., p. 68.

⁹ Ibid., pp. 70ff.

¹⁰ See Geertz, *The Interpretation of Cultures*, chap. 2. See also my *Value and Justification*, pp. 383ff.

¹¹ In her account of moral progress, Elizabeth Anderson stresses both overcoming moral biases (what I have called “climbing the ladder of impartiality”), but also experiments that help us better understand social dynamics and how to improve cooperative endeavors. See, for example, her “Social Movements, Experiments in Living, and Moral Progress.”

deep evaluative disputes. If differences in categorizations go as “deep” into a person’s perspective as evaluative standards, then the (controversial) claim that we do not truly disagree in evaluative standards is not much help to the Millian-Deweyian proponent of progress, for the remaining disagreements go just as deep.

At this point, the dialectic is unlikely to be advanced by looking at additional specific cases, for the dispute is about how we are to interpret the cases. The Millian-Deweyian and her critic can accept the cross-cultural data about differences in categorizations, but it remains open whether these are simply differences between pre-scientific “superstition” and scientific culture. The deeper problem, I propose, is an objectionable understanding of the basis of categorizations, and even their role in science itself. The rejection of categorization as a significant and deep part of a perspective would certainly make sense if something along the lines of Naïve Naturalism held:

1. The world is ultimately constituted by all the facts there are (the complete set of facts);
2. Science is the progressive uncovering of increasingly accurate representations of these facts; each new scientific discovery presents a set of facts that strictly dominates earlier versions (the later includes all that is correct in the earlier version, and adds to it).
3. Sound reasoners who (a) accept the best current science describing the world and (b) share the same evaluative standards will agree in their moral judgments.

On Naïve Naturalism, scientific progress secures, as it were, ever-better resolutions of the facts that compose the world. It is as if blurry and confused pictures become increasingly better representations of what *is*. If Naïve Naturalism held, disputes about categorizations would be disputes that science ultimately resolves, and a reasonable person should embrace science’s current picture, since it is more accurate than previous ones. Perspectives that accept the best current science thus would have the best available picture of the world, and so would not dispute the proper description of the natural or social world—only about the evaluation of the facts.¹² All sensible moral disputes would have to derive from their evaluative standards. If we are also in agreement about them, progress will lead to increasing moral agreement.

Naïve Naturalism—though seldom stated quite this baldly—is, I would venture, at the root of a great deal of secular moral and political theory. After all, if how we conceive of the world is really uncontroversial, isn’t it only evaluative standards that could cause deep reasonable disagreements? I believe this radically misconstrues science, and this because it fundamentally misconstrues the relation between discovery, theory, and categorization. Common-sense science accepts something along the lines of Naïve Naturalism, and within a shared paradigm that agrees on basic categories, measuring devices, and testing procedures, it may well seem to inquirers that Naïve

¹² It appears to also suppose either (i) that social facts are irrelevant to moral judgment, (ii) they are themselves subject to Naïve Naturalism’s view of scientific discovery, or (iii) they are reducible to the natural facts subject to science. None seems plausible. Stanley Benn and I long ago explored the relation of social categorizations (social facts) and normative theory in our *Public and Private in Social Life*, chaps. 1 and 2. On the implausibility of (iii), see Emile Durkheim, *The Rules of Sociological Method*.

Naturalism tracks their assumptions. However, different scientific programs are often based on very different categorizations of the world, and so represent the world in fundamentally different ways. Kuhn observes:

After about 1630, for example, and particularly after the appearance of Descartes's immensely influential scientific writings, most physical scientists assumed that the universe was composed of microscopic corpuscles and that all natural phenomena could be explained in terms of corpuscular shape, size, motion, and interaction. That nest of commitments proved to be both metaphysical and methodological. As metaphysical, it told scientists what sorts of entities the universe did and did not contain: there was only shaped matter in motion. As methodological, it told them what ultimate laws and fundamental explanations must be like: laws must specify corpuscular motion and interaction, and explanation must reduce any given natural phenomenon to corpuscular action under these laws. More important still, the corpuscular conception of the universe told scientists what many of their research problems should be.¹³

On the categorizations of corpuscular science, for example, gravity was an “occult quality” with highly suspect metaphysics; because action-at-distance appealed to categorizations and laws inconsistent with corpuscular science, it was implausible.¹⁴ More generally, if representation *A* is based on categories inconsistent with those a research group employs, it will not find *A* plausible regardless of the evidence marshalled in favor of it. Scientific progress, we might say, is based on shared categorizations, but it does not produce them. And, of course, while fundamental disagreements about categorizations characterize the history of disputes in the natural sciences, they are endemic to the social sciences.¹⁵

One of Hayek's vastly underappreciated contributions was integrating into social philosophy a theory of the mind according to which an agent's fundamental categorizations depend on her life experiences and the patterns of phenomena she regularly confronts.¹⁶ As Hayek stressed, all categorization depends on ignoring some differences (which become non-differences) and focusing on others.¹⁷ At its most basic level, a categorization is a many-to-one mapping: members of a set $\{a, b, c, d, e\}$ are mapped onto categories *W* and *X*, such that $\{a, d, e\} \rightarrow W$ and $\{b, c\} \rightarrow X$.¹⁸ Note that under this perspective *b* and *c* are the same (they are *W*s): difference between them is thus lost (an embryo is just a clump of cells, like a tumor). Yet on another perspective, where $\{a, b, d\} \rightarrow Y$ and $\{e, c\} \rightarrow Z$, *b* and *c* are different (perhaps entirely different) phenomena (perhaps one has a soul and so is the height of creation). Alf's and Betty's categorization schemas may differ in three ways. (i) As in the preceding example, their categorizations may take the same underlying entities $\{a, b, c, d, e\}$ and group them

¹³ Kuhn, *The Structure of Scientific Revolutions*, p. 41.

¹⁴ Ibid., p. 105.

¹⁵ See my “Diversity in the Moral Sciences.”

¹⁶ Hayek, *The Sensory Order*.

¹⁷ Ibid., pp. 48–52.

¹⁸ See Page, *The Difference*, chap. 3. This is entirely consistent with there being one real world that we all inhabit, and we all recognize this to be the case.

together in different ways, such that for Alf $\{a, d, e\}$ are one thing and $\{b, c\}$ another; while for Betty $\{a, b, d\}$ are one thing and $\{e, c\}$ another. Note here that their perspectives agree on the basic (“atomic”) entities $\{a, b, c, d, e\}$ but categorize them in different ways. (ii) Another possibility is that one perspective may omit from its classifications entities that are captured in another.¹⁹ (iii) Lastly, the perspectives may disagree on the basic entities, as corpuscular and Newtonian science did.

In his analysis of economic change, Douglass North has stressed the importance of Hayek’s work and the “amazingly modern resonance in recent work in cognitive science.”²⁰ Inspired by Hayek, North argues that economic development and social change are deeply influenced by the categories and mental structures that we cultural humans develop—a claim endorsed by fieldworkers in economic development.²¹ On North’s view, economics has fundamentally misunderstood economic change and development because it has presupposed a view of agents with exogenous preferences and a naïve theory of representational beliefs. We can longer accept that “there is a single universally valid model of the world.”²²

§16 Autocatalytic Diversity

16.1. Beinhocker’s Tale of Two Tribes

Diversity of perspectives is critical to trade, for different perspectives bring to the market different tools and approaches that lead to mutual gains.¹ As we saw in the first Part, (§8.3), perhaps the fundamental features of human ultra-sociality are relations of reciprocity. Being both self-interested and cooperative, reciprocation and trade are the foundations of all human cooperation among non-kin, from LPA hunter-gatherers to the contemporary technology economy. It is our differences that drive exchange. One of the ironies of much traditional economic modeling is the assumption of essentially homogenous utility functions—without differences there would be no markets to model.² In the *Wealth of Nations*, Smith argued that the tendency to “truck,

¹⁹ In these first two differences between Alf and Betty, they may formally share what Page (*The Difference*, chap. 3) calls “one-to-one mappings of the world,” which involves having a unique name for each element $\{a, b, c, d, e\}$. In the last case, not only are the entities grouped differently, there is disagreement about what they are.

²⁰ North, *Understanding the Process of Economic Change*, p. 33. North develops at some length the sort of connectionist theory of the mind explored by Hayek. See also Bicchieri, and McNally. “Shrieking Sirens—Schemata, Scripts, and Social Norms.” On connectionism, see Gärdenfors, *Conceptual Spaces*, p. 41.

²¹ For example, a belief in witchcraft: “to find a beehive with honey in the woods is good luck; to find two beehives is very good luck; to find three is witchcraft.” Platteau, *Institutions, Norms and Economic Development*, p. 201. See also Bicchieri, *Norms in the Wild*, chap. 3.

²² Solow, “Economic History and Economics,” p. 330; North, *Understanding the Process of Economic Change*, p. 19.

¹ See Muldoon, Borgida, and Cuffaro, “The Conditions of Tolerance.”

² See Miller, *A Crude Look at the Whole*, pp. 9ff, chap. 4.

barter and exchange” is “common to all men, and is to be found in no other race of animals, which seems to know neither this nor any other species of contracts . . . Nobody ever saw a dog make a fair and deliberate exchange of one bone for another with another dog.”³ To this, Edwin Cannan quips, “It is by no means clear what object there could be in exchanging one bone for another.”⁴ Cannan’s quip is spot-on: no matter how much two agents may tend toward reciprocity, if they have identical goods and preferences, no exchange will occur.

While difference and reciprocity beyond kin have characterized human ultrasociality since the Late Pleistocene era, clearly something quite remarkable recently has occurred in the development of these reciprocal relations. Eric D. Beinhocker tells a tale of two tribes, the Yanomanö, a contemporary hunter-gatherer tribe living along the Orinoco River between Brazil and Venezuela, and the New Yorkers, a latte-drinking tribe on the banks of the Hudson. It is obvious that the New Yorkers are far wealthier: their average income in 2001 was about \$36,000, about 400 times greater than the Yanomanö’s estimated \$90. We are well acquainted with understanding differences in terms of wealth inequality. But this overall income differential pales compared to the diversity differentials of their economies.

Retailers have a measure, known as *stock-taking units*, or SKUs, that is used to count the number of products sold by their stores. For example, five types of blue jeans would be 5 SKUs. If one inventoried all the types of products and services in the Yanomanö economy, that is, the different number of stone axes, the number of types of food, and so on one would find that the total number of SKUs in the Yanomanö economy can probably be measured in the several hundreds, and at most in the thousands. The number of SKUs in the New York economy is not precisely known, but using a variety of data sources I very roughly estimate that it is on the order of 10^{10} (in other words, tens of billions). To put this enormous number in perspective, estimates of the total number of species on earth range from 10^6 to 10^8 .⁵

While the New Yorkers’ wealth is 400 times greater than that of the Yanomanö’s, the diversity of the products and services available to them is on the order of a hundred-million-fold.

16.2. Dividing Labor vs. Creating Functions

The basic agential relation, one of rule-governed reciprocity, on which all human ultrasociality depends, has grounded an explosion of goods and services in the contemporary Open Society. How? One of Smith’s great insights is that specialization and the division of labor are driven by the extent of the market.⁶ Within a small group,

³ Smith, *The Wealth of Nations*, vol. I, p. 17.

⁴ *Ibid.*, p. 17n.

⁵ Beinhocker, *The Origin of Wealth*, p. 9.

⁶ Smith, *The Wealth of Nations*, vol. I, pp. 21–5. Thus, as Smith predicts, the development of markets in Africa has been slowed by the dispersal of population and lack of transportation routes. Platteau, *Institutions, Norms and Economic Development*, pp. 42ff.

specialization cannot proceed far: since the ability to trade is limited by the number of those interested in what one has to offer, in small populations the viable specialization niches will be very broad and general. “Country workman are almost everywhere obliged to apply themselves to all the different branches of industry that have so much affinity to one another as to be employed about the same sort of materials. A country carpenter deals in every sort of work that is made from wood: a country smith in every sort of work that is made from iron.”⁷ As the market expands, new niches become viable. In a large market, making a specific type of wooden toy may be a viable niche—indeed, entirely new niches, such as a porters, may become a specialization.

Trade, then, depends on differences, and the greater the extent of the market, the finer-grained those differences. Smith conceived the extent of the market in terms of its size: the greater the population linked in a market, the more specialized niches become viable occupations. For Smith, adding population to a market (say, by expanding transportation networks) adds diversity because of the productive advantages of specialization. As we all know, his classic example was the pin factory: one worker, he estimated, could by himself make twenty pins a day, but by dividing the process into ten subtasks (straightening the wire, putting heads on, etc.), Smith estimated that the per worker output was almost five thousand pins a day.⁸ What is so misleading about this classic example, however, is that the division of labor simply divides up a known task into its component parts: the entire set of tasks was known, the division simply allowed us to separate them into different workers with complementary functions. If one thinks of the division of labor in terms of distributing something like a fixed stock of “available” jobs,⁹ dividing labor can be modeled as a coordination game, as in Display II.1.

		Betty		
		Specialize Roofs	Specialize Structure	Don't Specialize
Alf	Specialize Roofs	3 rd 3 rd	1 st * 3 rd	2 nd 3 rd
	Specialize Structure	1 st * 1 st	3 rd 3 rd	2 nd 1
	Don't Specialize	3 rd 2 nd	3 rd 2 nd	2 nd * 2 nd

Display II.1. Dividing functions as coordination.

In this game, Alf and Betty can each build an acceptable house alone (the second-best outcome). If Alf and Betty specialize, they get very adept at their particular task, but their skill for the other task becomes degraded. If one specializes on the roof and

⁷ Smith, *The Wealth of Nations*, vol. I, p. 21.
⁸ Ibid., pp. 8–9.
⁹ This supposition underlies much of the analysis of O'Connor, *The Origins of Unfairness*, p. 70.

the other on the structure and they trade services, they get better houses (their 1st ranked outcome). However, if one specializes and the other doesn't, the one who doesn't specialize builds the house alone (getting their second choice), but the other has degraded their skill on half the process and overall, the house suffers (so they get their worst outcome). There are three Nash equilibria (marked by *)—two in which they specialize on different tasks, and one in which neither specializes. Given this, the parties have an interest in achieving a correlated equilibrium that instructs one to roof while the other frames the house.¹⁰ Smith's pin factory has such a correlating mechanism: the factory manager assigns tasks. A critical function of hierarchy is to identify a correlator, such that its decision directs people to their role. (We thus see that not all hierarchy is domination.)¹¹ Plato's *Republic* generalizes this device to social and political life in general. Alternatively, the players can use other cues, such as one's parent's occupation, or, as O'Connor stresses, ethnic or gender markers ("Women don't do roofs!").

The critical point, however, is that this model is truly enlightening when there are stable complementary jobs to be filled, and it is critical that they each be filled in some sort of balanced way. In that case, we are indeed confronted with a coordination-allocation problem. But the point of Beinhocker's tale is that new tasks and jobs—we might think of these as *niches*—are being constantly created, and successful niche construction depends on attracting others in complementary niches, and outcompeting similar niches for their interaction. Not only do the New Yorkers perform tasks and make trades that are unimaginable to the Yanomanö (including in financial instruments that are unimaginable to most of us, and perhaps barely to them), but in this constantly changing environment, even existing jobs are constantly adjusting to their new environment. Think of the tremendous changes that have occurred even in "stable" niches such as doctor or teacher: here there is competition within existing functions to revise them in ways that attract others, and outcompete traditional specifications of the function. To understand the hundred-million-fold explosion in diversity, we need to know how we have come to regularly exchange things and services that were until recently unimaginable. If that is so, a model about allocating fixed positions will be distinctly unhelpful.

16.3. McCloskey's Open Society Thesis

How, then, has this tremendous explosion of the diversity in the market occurred? The obvious proposal is, again, Popperian: the Open Society is a marketplace in new and original ideas, which spurs constant innovation.¹² If we wish to know why the division of labor has resulted in hitherto unimaginable types of labor being divided in hitherto unimaginable types of ways, we need to know what sort of society induces

¹⁰ In such correlation equilibria, players condition their play on some cue, such that when that cue is present, one will specialize in roofs and the other on structure. On correlated equilibrium, see Vanderschraaf, *Strategic Justice*, pp. 59ff. O'Connor's analysis employs correlated equilibria, though she is reluctant to appeal to the idea, seeing hers as a novel solution. *The Origins of Unfairness*, pp. 55ff.

¹¹ See Coase, "The Nature of the Firm"; Williamson, *The Economic Institutions of Capitalism*.

¹² McCloskey, *Bourgeois Equality*, p. 40.

innovative thought and products. This is the crux of Deirdre Nansen McCloskey's recent encyclopedic study of the growth of wealth in western Europe in the last two centuries. On her analysis, the explosive growth in the market and the accompanying human betterment was driven by a transformation in ideas, both moral and market-tested innovations. The core of her thesis is that in the last two centuries (further back for the Dutch and the English), "there was a sharp rise in society's receptiveness to improvers. It was social memes, socially inherited ideas, that changed ... Liberty and dignity meant that the society was receptive to trade-tested betterment ..." ¹³ For McCloskey there was, then, first a change in thinking about morality and virtue, such that the bourgeois virtues of honesty, industry, and respect for commoners supplanted aristocratic ways of relating. In terms of Part I, she argues that there was a subordinate rebellion, which decreased the social and political power of those at the apex of the hierarchy in favor of more egalitarian cooperation. The moral transformation produced by this rebellion opened society up to widespread innovative thinking—innovation tested for its worth in terms of market success. And thus the explosion of new services and products.

McCloskey's analysis reminds us of the importance of liberal norms in the evolution of a highly diverse Open Society. In evaluating her account, however, the critical issue is precisely what is meant by a culture of innovation. By definition, if we wish to explain how new diversity arises, our explanation must be about some sort of innovation. Culture proceeds through innovations—otherwise there would not be cultural evolution. Yet, we have seen that a culture of bold risk-takers seems *diversity reducing*: a culture that focuses on innovation is, surprisingly, more apt to lose than gain diversity. The key to maintaining diversity is to maintain already successful innovations, and for that we need a bevy of high-fidelity copiers (§11.2). But if that is so, how can we explain the explosive diversity of the Open Society?

16.4. The Exploration of the Adjacent Possible

Stuart Kauffman has recently stressed the similarity of the explosion of diversity in the biosphere and the economy by the exploration of the adjacent possible. "We seem to make our worlds and thereby make room for one another. Each ... makes even more opportunities for others in its adjacent possible niches or rooms. The adjacent possible niches, like worms coming to live in swim bladders, *explode faster than the number of occupants* who, by existing, create those very adjacent possible niches." ¹⁴ To better understand the importance of the adjacent possible in the evolution of diversity, let us consider the supply side and the demand side of innovation.

¹³ Ibid., p. 472.

¹⁴ Kauffman, *A World beyond Physics*, pp. 106–7. Emphasis added. Compare this to Smith's population-based analysis of the development of the market (§16.2). On Kauffman's analysis, diversity outstrips the growth in population. See also Holland, *Complexity*, chap. 7.

The Supply Side

Innovations and ideas rarely emerge newborn out of the head of a “man of genius,” but are almost always variations of what already exists.¹⁵ The amplifier circuit, for example, was a development of the triode vacuum tube, which made the oscillator possible, which eventually led to radio broadcasting.¹⁶ The same tale can be told again and again. At each step along the way, the ability to make a move was based on the previous technological moves already made, on which the latest mover capitalized. This *directed tinkering* model of innovation is critical to much technological change.¹⁷ This, says Ridley, “is the story of technology, from the Stone Age to the present day, on all continents, wherever you look, technology proceeds in a stately way from each tool to the next, rarely leapfrogging or side-stepping.”¹⁸ However, as Arthur stresses, not all such change is so stately: truly novel changes emerge by a process of evolution by combination—inventive recombination of different technologies.¹⁹ Even here, innovation depends on previous innovations; as innovations increase, their possible modifications and new uses become ever greater. Hence technological growth is *autocatalytic*: the more innovations there are, the greater the space of the adjacent possible to the present, which allows for even more innovations, once again expanding the space of the adjacent possible. Because innovation is autocatalytic in this way, one of the best predictors of how much innovation there will be is its present degree of diversity, for that expands the space of the adjacent possible. Diversity begets diversity.²⁰

Note that in these cases innovation itself requires copying. An innovator must know precisely how to produce a triode vacuum tube if it is to be the basis of an oscillator. Movement into the adjacent possible technology is thus a delicate mix of high-fidelity copying of the current technology and explorative tinkering with of it. High-fidelity copying is critical at both the start and end of the innovation process. Unless the innovators have copied the present technology accurately, success is largely random;²¹ unless the innovation is itself successfully copied, it is of no avail.²²

While many might be willing to grant this autocatalytic account for technology, it is apt to be resisted concerning theoretical intellectual achievements. Here, the truly exceptional solitary genius may still seem dominant. But this too is dubious. We know that many of the great intellectual breakthroughs—the calculus, evolutionary theory itself, the double helix, the theory of special relativity, and, more modestly, marginal utility—were either simultaneously discovered or the evidence indicates close work by others at the time of the breakthrough.²³ This is certainly not to belittle Newton,

¹⁵ Arthur, *Complexity and the Economy*, p. 141.

¹⁶ *Ibid.*, p. 141. See also Beinhocker, *The Origins of Wealth*, pp. 246–8.

¹⁷ *Ibid.*, pp. 249–53.

¹⁸ Ridley, *The Evolution of Everything*, p. 125.

¹⁹ Arthur, *Complexity and the Economy*, pp. vii, 120.

²⁰ *Ibid.*, p. 146. More generally, “[r]ecombining mechanisms implicitly leverage performance criteria of new agents and strategies because they draw parts of the new agent or strategy from those that are already succeeding.” Axelrod and Cohen, *Harnessing Complexity*, p. 43.

²¹ Which is not to say that random changes cannot produce functional change. As we know from genetic evolution, that is possible, but the pace of adaptive change is slower than in more directed change.

²² Beinhocker, *The Origins of Wealth*, pp. 253–4.

²³ This is just a sampling. See Ridley, *The Evolution of Everything*, pp. 118ff.

Darwin, Watson, Einstein, or Jevons, but rather to stress that intellectual innovation overwhelmingly occurs within the adjacent possible. It can take a genius to see the next move into the adjacent possible, but at that point other geniuses are apt to be working on this move. Again, I do not wish to be dismissive, as if the adjacent possible space is simply there for all with eyes to see: creativity is critically a matter of reconceptualizing the adjacent possible. And that is why employing different categorizations and different similarity relations is so important: you see possibilities I do not (§15.2). Similarity judgments depend on categorizations—only once a perspective categorizes *X* and *Y* can it determine how similar they are. A system of categories reveals meaningful structure in the world, and a critical component of this structure is relations of similarity. Once one has categorized a phenomenon, relations of similarity are generated.²⁴ Now it is the differences in similarity judgments induced by difference in perspectives that makes discovery and innovation creative and often unexpected, for differences in similarity imply differences in what is the adjacent possible, and so what are possible “next moves” in discovery. So once again, diversity (now of categorizations and similarity spaces) begets further, diverse, ideas. As always, such innovation still requires high-fidelity learning of the current view, combined with high-fidelity copying of the innovation. Thus—in contrast to the traditional idea of creative discovery as great leaps into the unknown by the “man of genius”—the exploration of the adjacent possible by diverse perspectives can explain autocatalytic innovation within the context of cultural learning and conformity.

As Kauffman observes, the adjacent possible drives the development of the market itself:

As new goods and services and production capacities come into existence, they provide the growing contexts into which yet more new goods and services and production capacities can follow as their complements or substitutes. An economy with a high diversity of goods, services, and production functions is rather like a garage full of “stuff” rather than a clean garage. It is easier to jury-rig in the garage full of stuff, and it is easier to invent new goods and services and production functions in an economy already full of such stuff. But the new goods, services, and production functions only make the “garage” more full of stuff; *thus, amazingly, the economy grows its own adjacent possible and augments that very growth as the growth occurs. The process is broadly self-accelerating.*²⁵

The growth of diversity is self-accelerating because the size of the possible exponentially increases with the dimensions of the attributes being tinkered with.²⁶ If all I have in my garage are ping-pong balls, no matter how filled it is, there are few distinct features to play with (say, size, ability to bounce, color): if I add some strings and hooks I can make

²⁴ For example, suppose, one’s category of “punishment” is characterized simply by incarceration due to criminal behavior. This categorization also implies similarity judgments about punishments: those that involve similar length of incarceration are similar punishments, while those that involve very different lengths are very dissimilar punishments. I consider similarity judgments in considerable depth in *The Tyranny of the Ideal*.

²⁵ Kauffman, *World beyond Physics*, p. 137. Emphasis added.

²⁶ See Page, *Diversity and Complexity*, pp. 13–31.

a decoration for my Christmas tree; add some glitter and lights, I can use them as decorations for my disco party. Add some nitroglycerine and they can turn deadly.

The Demand Side

Thus far I have been stressing that, as diversity expands, the space of the adjacent possible expands at an increasing rate. On Kauffman's metaphor, as our garage fills up with more and more stuff (as mine always is), the things one can do with it vastly increase. This might be deemed a "supply side" analysis, as it focuses on the supply of current niches that allows for new adjacent possible ones. Arthur nicely shows the interrelation between the supply side and demand side of the process:

Growth in coevolutionary diversity can be seen in the economy in the way specialized products and processes within the computer industry have proliferated in the last two decades. As modern microprocessors came into existence, they created niches for devices such as memory systems, screen monitors, and bus interfaces that could be connected with them to form useful hardware-computing devices. These, in turn, created a *need, or niche*, for new operating system software and programming languages, and for software applications. The existence of such hardware and software, in turn, made possible desktop publishing, computer-aided design and manufacturing, electronic mail, shared computer networks, and so on. This created niches for laser printers, engineering-design software and hardware, network servers, modems, and transmission systems. *These new devices, in turn, called forth further new microprocessors and system software to drive them.* And so, in about two decades, the computer industry has undergone an explosive increase in diversity: from a small number of devices and software to a very large number, as new devices make possible further new devices, and new software products make possible new functions for computers, and these, *in turn, call forth further new devices and new software.*²⁷

A new niche can "call forth" work on the adjacent possible, inducing an exploration that will better solve the problems, or exploit the capabilities, of the current technology. This demand-side exploration is functional: given the current niche and its technological development, something is required that does a job that is not being done. The next step in the development of a technology, like the next step in the evolution of an organism, depends on solving a functional problem. But, as Kauffman stresses, functions are inherently parts of adaptive wholes: an innovation is only functional if it fits into a system that is adaptive, and the innovation assists in achieving and maintaining that adaptiveness.²⁸ The demand side, then, is critical insofar as it manifests a function oriented search of the adjacent possible.

²⁷ Arthur, *Complexity and the Economy*, p. 169. Emphasis added,

²⁸ Kauffman, *World beyond Physics*, chap. 2. In different fields, adaptiveness can be defined in different ways. In the market it is utility to consumers. McCloskey's gloss is that the market is a way for testing the adaptiveness (usefulness) of innovations. See §14.2 in this volume for her idea of "trade-tested betterment."

Kauffman defends what we might call the *Principle of the Indefiniteness of Functional Searches*. He takes the example of a simple tool (a functional item) such as a screwdriver and advances two claims: “(1) no rule-following procedure, or algorithm, can list all the uses of a screwdriver; and (2) no algorithm can list the next new use of the screwdriver!”²⁹ For even such a simple tool, one cannot exhaustively list all the things that might be done with it. “Please list for me all the uses of a screwdriver in, say, New York, in 2017. Well, go ahead: screw in a screw, open a can of paint, scrape putty from a window, stab someone, display as an *objet d’art*, scratch your back, wedge the door open, prop a window open, jam a door closed, tie to a stick and spear a fish, rent the spear out at 5% of the local catch, and so on.”³⁰ There can be no predictive theory of the evolution of the screwdriver since the possibility space of functional uses cannot itself be determined. “We do not know the sample space of biological evolution, whose becoming is therefore not a machine.”³¹ This inability to predict is not because the excessive magnitude of the data required for such a prediction (which, some might hope, will eventually be overcome), but because the next innovation in evolution depends on hitting upon a new function, which is not a defined possibility space.

Still, one might wonder—if innovation occurs in the adjacent possible, can’t we say that it is more likely to occur in the “nearest” part of that space? Recall our discussion of perspectives and categorizations (§§15.2–3). Different perspectives differ in their categorizations, hence their notions of what is close to the current “technology” also differ. Thus here, too, there is diversity about what is adjacent.³² And again, it is often a productive diversity, as it effectively expands the space of the adjacent possible, spurring innovations—and, of course, new diversity. In the evolution of technology, what constitutes a functional development depends on the varying perspectives on the adjacent possible and the understandings of the pressing functional needs of existing adaptive units and how they are categorized. Add a new perspective and its categorizations, and the possibility space expands.

16.5. Autocatalytic Diversity and the Open Society

Let me try to bring together the somewhat involved discussions of the first three sections of this Part. In section 14, I expressed considerable skepticism that traditional Millian-inspired progressivist defenses of free inquiry adequately support the depth and range of diversity constitutive of the Open Society. This then led us to an inquiry into the possible range and depth of diversity. Section 15 explored the idea of diverse perspectives, with the main aim of stressing the sources of diversity, and the importance of often-neglected dimensions of diversity, such as differences in categorizations, which ill-fit the classical “experimental” Millian model. Whereas normal science—the model that usually inspires the Millian-Popperian account—assumes shared categorizations, I explored a deeper diversity for which there is considerable

²⁹ Kauffman, *World beyond Physics*, p. 119.

³⁰ *Ibid.*, p. 118.

³¹ *Ibid.* This is also Hayek’s view. “The Theory of Complex Phenomena,” p. 274.

³² This idea was at the heart of *The Tyranny of the Ideal*, esp. chap. 2.

evidence. Section 16 has applied this analysis to reciprocity, specialization, and the division of labor.

On this analysis, then, in a fundamental way the Open Society is continuous with all human social order, being based on the benefits of reciprocity. But, then again, it is strikingly noncontinuous in the depth and range of diversity that it sustains and generates. Thus the core question of Section 16: What has driven this explosive growth in diversity? Smith points to the extent of the market, and that was surely an important part of the story, but size is not everything: the Chinese population in 1850 was about a hundred million greater than Europe's, yet the latter had a much more developed division of labor. McCloskey thinks the critical development was bourgeois ethics, and the way it facilitated new ideas and innovations to be tested in the market. This too is surely part of the story, but stressing innovation and new ideas in commercial culture runs smack into our analysis of *culture*, which stresses the human facility at copying, and the importance of imitation in maintaining a diverse society. "The mechanism of copying the interaction patterns of other agents passes along vital social knowledge and allows an agent to adapt, without requiring an explicit understanding of very complex social systems."³³ The key to reconciling innovation and imitation, I have argued, is appreciation that diversity and innovation rarely grow by wild leaps of radical creativity, but by tinkering, recombining and reconceptualizing the adjacent possible. Such growth requires accurate knowledge of the current culture, and a facility at seeing how complementarities, or alternatives, can be developed. Yet, just what possibilities are "adjacent" is fundamentally a matter of categorizations and perspectives. Two innovators may both successfully copy a current innovation, yet the way they categorize the world may lead them to explore vastly different "adjacent possibles." Thus we arrive at a fundamental result, reconciling the critical roles of high-fidelity copying and searching the adjacent possible while understanding how innovators take surprising moves.

Lastly, and most importantly, the growth of diversity is autocatalytic. As culture and the market become more diverse, perspectives multiply and so the space of the adjacent possible explodes: innovation creates even more possibilities, which when realized further expand the adjacent possibles. As McCloskey reminds us, this is not inevitable: hostile cultural norms and authoritarian regulation can impede searching and developing the adjacent possible. But when there is searching, it inevitably produces the type of runaway diversity that characterizes the Open Society.

§17 Diversity and Complexity

17.1. Interactive Complexity

Diversity need not lead to complexity. Surnames in New York City display great diversity but not complexity. Autocatalytic diversity, however, does lead to complexity.

³³ Axelrod and Cohen, *Harnessing Complexity*, p. 89.

Once again, the clearest examples come from the development of technology—the original jet engine had one moving part (a compressor-turbine), while current engines have over 22,000.¹ Efficiency was enhanced by adding a number of compressors linked into a system, then adding variable vanes to regulate air intake, then airflow systems to reduce overheating, and on and on. What started as a remarkably simple system is now a highly complex one. The complexity arose from innovations in the space of the adjacent possible, which led to increasingly complex networks of innovations. Each innovation made for a more complex jet engine, which then provided the basis of the next innovation. But to be successful, the next innovation had to work with the existing set, and so the functioning of the innovation was critically linked into that functional network we call the contemporary jet engine. This network went well beyond simply the internal functioning of the engine to include the development of alloys and computer systems linking the jet engine to the rest of the aircraft and beyond. As with all explorations of the adjacent possible, the development was by no means predetermined, but was a path-dependent process depending on a long series of explorations and tinkering in a high-dimensional environment, with innovators acting on different notions of the adjacent possible.

Complexity arises when the behavior of heterogeneous elements are connected such that the functioning of one part affects many others.² In a jet engine or a biological organism, this means that a variation in the behavior of one element feeds back to the behavior of others, which then affects others, and so on. Sometimes these radiating consequences can be halted (as in a circuit breaker) or dampened (as in contemporary stock markets, which automatically pause trading for a short length of time to slow down cascading positive feedbacks).³ But complex systems cannot function unless their elements interact.

Today's societies are extraordinarily complex because high levels of diversity are conjoined to high *reflexivity* in the behavior of each agent. Beinhocker calls this “interactive complexity.”⁴ Each individual is acting on her perspective (§15) where this requires that she is constantly reacting to the choices of others about how to adaptively act on theirs. As new patterns of interactions are formed, individuals react to them, while others react to those changes. “Complexity . . . asks how individual behaviors might react to the pattern they together create, and how that pattern would alter itself as a result.”⁵ In our contemporary complex society, then, networked individuals create patterns through their interactions and then react to those very patterns, constantly producing new patterns.⁶ As Donald Saari concludes:

even the simple models from introductory courses in economics can exhibit dynamical behavior far more complex than anything found in classical physics or biology. In fact, all kinds of complicated dynamics (e.g., involving topological entropy, strange

¹ Arthur, *Complexity in the Economy*, chap. 9. See also Laland, *Darwin's Unfinished Symphony*, pp. 10–1.

² Axelrod and Cohen, *Harnessing Complexity*, p. 7.

³ Miller, *A Crude Look at the Whole*, chap. 3. These words were written a day after such a breaker halted trading on the New York Stock Exchange in March 2020.

⁴ Beinhocker “Reflexivity, Complexity, and the Nature of Social Science,” p. 332.

⁵ Arthur, *Complexity and the Economy*, chap. 1.

⁶ Ibid.

attractors, and even conditions yet to be found) already arise in elementary models that only describe how people exchange goods (a pure exchange model). Instead of being an anomaly, the mathematical source of this complexity is so common to the social sciences that I suspect it highlights a general problem plaguing these areas.⁷

The hidden complexity of social science derives from aggregation out of the unlimited variety of preferences, “preferences that define a sufficiently large dimensional domain that, when aggregated, can generate all imaginable forms of pathological behavior.”⁸

As Saari points out, given heterogeneity (and reflexivity), extreme complexity can arise even in simple economic models. But human complexity is far more extreme than standard economic models would indicate. In most neoclassical economic models, the economy is characterized by strong negative feedbacks.⁹ The fundamental assumptions of decreasing marginal utility and decreasing marginal returns suppose that the more an individual engages in an activity, and the more other individuals enter a market, the less apt she is to continue that activity or expand in that market. That is why neoclassical economics can so often posit a unique equilibrium.¹⁰ No matter where we begin, given strong negative feedbacks, we should, theoretically, end up at a unique equilibrium. However, contemporary economies and societies are characterized by strong positive feedback forces: the more a society depends on networking (such as operating systems, social media, business communications),¹¹ the more positive returns dominate.¹² With positive returns, the more people who engage in an activity, the more attractive it is for others to join in. In systems with strong positive return dynamics, path-dependency abounds: the state of the system at time $t + 1$ critically depends on its state at t . Because of this, rather than gravitating to a unique equilibrium, multiple equilibria abound; where you end up critically depends on the order in which events occur.¹³

17.2. The Threat of Chaos

To put the problem in somewhat more prosaic terms, as societies become more diverse the elements becomes increasingly interconnected. As in the evolution of the jet engine,¹⁴

⁷ Donald Saari, “Mathematical Complexity of Simple Economics,” p. 222. Strange attractors are related to chaos theory; see Smith, *Explaining Chaos*, pp. 142–6.

⁸ Saari, “Mathematical Complexity of Simple Economics,” p. 229.

⁹ Interestingly, Hayek also stressed the importance of negative feedbacks. See Lewis, “Purposeful Behaviour, Expectations, and the Mirage of Social Justice.”

¹⁰ See Beinhocker, *The Origin of Wealth*, chap. 2.

¹¹ As well as fashions and following trends.

¹² Positive returns are by no means unique to the contemporary world. Arthur gives the example of the familiar analog twelve-hour “clockwise” clock: in the fifteenth century there were clocks that went anti-clockwise, including the twenty-four-hour clock in a Florence cathedral. The selection of the familiar clock was path dependent: once most people began using the familiar twelve-hour clock, others had increased incentive to also adopt it. See Arthur, *Increasing Returns and Path Dependence in the Economy*, p. 2 (more generally chaps. 3, 8, 10).

¹³ See *ibid.* See also Holland, *Complexity*; Miller and Page, *Complex Adaptive Systems*.

¹⁴ Arthur, *Complexity in the Economy*, chap. 9. See also Laland, *Darwin’s Unfinished Symphony*, pp. 10–11.

what once were distinct systems (jets engines and vacuum tubes) grew together as jet engines became more complicated and vacuum tube technology evolved into microchips. As the economy evolves (in this sense), it is constantly becoming increasingly interconnected and complex. To be sure, this ever-increasing complexity is punctuated by periodic disruptions. Just because the economy is so interconnected when some technologies or practices are displaced by innovations, the disruptive effects are apt to be far-reaching—and surprising. As Schumpeter stressed, complex economies are subject to the “gales” of creative destruction: entire areas of the economy may ultimately be wiped out when there is a major technological change. For example, the invention of the automobile wiped out a wide variety of trades, from carriage makers to horse breeders, grooms, and stables, to dung sweepers.¹⁵ The effects of these disruptive gales may be astonishingly wide, changing far-flung and surprising parts of the economic and social system. The automobile gale changed the face of modern societies, including the structure of cities, adolescent and female independence, family relations, housing, sex roles, internal migration patterns, and on and on.¹⁶ With both consumer and social technologies (banking, the stock market, universities), simple technologies are assembled and combined, modified with an eye to increasing functionality and solving new problems, thus producing ever-greater complexity. Of course, compared to human society a jet engine is astoundingly simple: experts (though perhaps now only in highly trained teams) continue to know how it works. But like technology, we have constantly rendered our institutions more complex, and more intimately tied to the rest of our social system.

Diversity generates ever more diversity, which generates complex interconnected systems. Why, though, does all of this generate functional systems—diverse systems that maintain their integrity and, in some broad sense, work? There is no guarantee that ever-increasing diversity leads to ever-increasing complexity, as opposed to system collapse in the face of conflict and tensions. Kauffman’s garage may be a treasure house of diversity, but it does not form an adaptive system. Complexity requires interconnections, but just as diversity with too few connections is simply a pile of stuff, Kauffman has also showed that with too many connections, the system becomes chaotic. In what is known as an *NK* model of complexity, we suppose *N* elements, with *K* being the average number of connections of each element with the others. Even if *N* is very high, if *K* = 0 we only have a pile of diverse stuff (Kauffman’s garage). But when *N* is high and *K* approaches *N* – 1 (each element’s behavior is linked to every other element’s behavior), chaos results because a change in the performance of any single element changes the performance of all the others.¹⁷ Each change essentially produces an entirely new system state, so the system randomly moves around the possibility

¹⁵ Schumpeter, *Capitalism, Socialism and Democracy*, pp. 81–7; Arthur, *Complexity in the Economy*, p. 141.

¹⁶ In 1893 an American newspaper surveyed seventy-four prominent American men and women, asking them for their predictions about life in 1993. “What is remarkable is not the omission of nuclear weapons or microelectronics; many prominent public figures foresaw both new weapons and new forms of electrically powered global communication. The striking oversight was motoring. Nobody understood the chain of technological, commercial, social and political events that would surround the internal combustion engine vehicle which Karl Benz has patented in 1886.” Tenner, *Why Things Bite Back*, p. x.

¹⁷ Kaufman, *The Origins of Order*, pp. 45ff. See further §§23.2, 27.1 in this volume.

space. Imagine a system with 100 on/off switches; turning switch 57 from on to off affects every switch, and each of them in turn changes every other switch: flipping a single switch moves the entire system to a radically different state. There is, then, no such thing as tinkering in such a system. Kauffman calls this “a complexity catastrophe.”¹⁸ Under such conditions the system is unlikely to maintain itself, since after a series of changes some system states are likely to be ones where the values of most of the elements are miserable.

17.3. The First Solution: Macro Selection

Hence the question presses: How can this ever-more complex system maintain its functionality rather than collapse into chaos? Why is there organized, as opposed to disorganized, complexity?¹⁹ Recently, David Sloan Wilson has forcefully argued that functionality can only be maintained if society is subject to adaptive pressures at the system level. We can treat this, I believe, as a macro-selection account (§11.4): “[f]rom an evolutionary perspective ... only when a society is a unit of selection ... does it function well as a unit.”²⁰ The critical claim is that a complex social order will maintain its cooperative functionality only if, at the societal level, forces are constantly selecting more over less functional social orders. This is very much in the spirit of Hayek’s group-focused theory that we examined in section 11.4.

Macro selection offers an attractive hypothesis as to why complex wholes maintain their adaptiveness, preventing complexity from degenerating into chaos. Herbert Simon has argued that evolutionary pressures dampen complexity by inducing decomposability.²¹ If a change in one element produced changes in all the others, a species could not climb an evolutionary gradient, becoming increasingly fit.²² If, say, every change in an organism’s ability to run faster affected all its other traits, it would be in a state of constant instability: marginally increased speed might seriously compromise a host of other traits. Decomposability, then, makes the system less complex qua less tightly interconnected. Suppose we have a set of elements, $\{e_1 \dots e_{25}\}$ with subsets $\{e_1 \dots e_{10}\}$, $\{e_{11} \dots e_{15}\}$, $\{e_{16} \dots e_{25}\}$, where the elements within each subset are highly interconnected, but the connections between the subsets are modest.²³ At a limit, each subset could be a module that could connect with the others such that regardless of the changes that occurred within in it, it could be “plugged into” the others

¹⁸ Ibid., p. 52. See also McKelvey, “Avoiding Complexity Catastrophe in Coevolutionary Pockets,” esp. pp. 301–2.

¹⁹ Miller and Page, *Complex Adaptive Systems*, pp. 48–53. See also Jacobs, *The Death and Life of Great American Cities*, chap. 22.

²⁰ Wilson, “Two Meanings of Complex Adaptive Systems,” p. 44.

²¹ Simon, *The Sciences of the Artificial*, pp. 183–216. For an excellent analysis, see D’Agostino, “From the Organization to the Division of Cognitive Labor.”

²² Other models indicate that high dimensionality may be an asset in evolving more adaptive organisms. See Gavrilits, “High-Dimensional Fitness Landscapes and Speciation.”

²³ There is, perhaps, an “interface” that allows the outputs of one subsystem to feed into another (thus there is truly an overall system, not two independent systems) in such a way that variations of the connections within a subsystem do not too sensitively affect its output into another. See D’Agostino, “From the Organization to the Division of Cognitive Labor.”

without inducing too many changes in them.²⁴ In this case we could optimize within each subset and then assemble the results, maintaining the optimization within each set. Decomposability is a windbreak on the gales of creative destruction, halting the systematic effects of many changes in a subsystem.

This analysis supposes that the adaptive pressure is at the group or societal level: strong pressures for group or social functionality ensure that successful systems control complexity to attain functionality. On this essentially macro-selection account, the social order of the Open Society is *manageably* complex just because *too* complex orders (i.e., those that were insufficiently decomposable) have been selected against in competition with other societies. Hayek's prescience is striking. Not only did he propose an account of social and economic orders as complex systems, but he recognized that such orders would be functional when the order was adaptive at the social level, with this being determined by inter-group competition. Wilson's "new" and controversial claim follows Hayek's by over half a century. Once again, we see the centrality of Hayek's "group selection" account to his overall account of the social order.

Alas, it does not seem that, regarding the contemporary "Great" or Open Society, the Hayek-Wilson stress on macro selection is well-grounded. To be sure, it is plausible to hold, as many do, that such macro social-selection pressures have been great during some epochs. As we saw in Part I (§11.4), in the Late Pleistocene era human groups were subject to severe selection pressures, and those orders that were less intensely cooperative were probably eliminated.²⁵ A similar case can be made for eras of intense warfare.²⁶ Such eras plausibly produced our cooperative nature. It is, I would venture, far less plausible to think that most societies are subject to sufficiently strong pressures of this sort today. To be sure, some societies do look severely dysfunctional and the expected mass emigration has occurred. But most social orders have become, as societies, sufficiently well-off that they can withstand competition with other societies without great adjustments.²⁷ For the same reasons that group selection pressures were so strong in the Late Pleistocene era—harsh, quickly changing climate, numerous socially distinct groups near the margin of viability—we would expect that our modern era, characterized by the absence of these features, would have greatly mitigated macro-selection pressures. One simply cannot invoke the cut-throat group competition in the Late Pleistocene era to ground the importance of group-level selection, and then say it proceeds today in a drastically less competitive environment. We must always remember that on a multilevel selection account, GROUP-level pressures must be strong enough to overcome INDIVIDUAL-level selection: it is not enough to simply say there are "significant" GROUP-level pressures—they must be reasonably

²⁴ In this respect, recall Rawls's claim that a public conception of justice is a module that fits into multiple comprehensive conceptions. *Political Liberalism*, p. 12. Rawls is claiming that people's normative systems are decomposable.

²⁵ See Bowles and Herbert Gintis, *A Cooperative Species*, chap. 6; Richerson and Boyd, *Not by Genes Alone*, pp. 224–9; Richerson and Boyd, "Rethinking Paleoanthropology: A World Queerer than We Supposed."

²⁶ Turchin, *Ultrasociety*.

²⁷ Compare here Schumpeter's analysis of late capitalism, in which firms have grown sufficiently large to weather the gales of creative destruction, and so competition and innovation slow. *Capitalism, Socialism and Democracy*, pp. 87ff.

strong.²⁸ Moreover, by its very nature, Hayek's Great Society—an expansive transnational network of rule-based cooperation—blurs the boundaries between groups on which GROUP selection depends. As complexity spreads and a global order arises, encompassing previously relatively distinct groups, it becomes increasingly obscure what we mean by “a society” competing against others. Models of group selection flounder when groups have highly permeable boundaries with low competition among them. Although some investigators see strong group (or macro-) selection pressures continuing today,²⁹ this seems, at best, highly conjectural.

17.4. The Second Solution: Self-Organization

The obvious alternative is a second type of complex adaptive system identified (and criticized) by Wilson: adaptation via adjustment by each actor to the previous and anticipated actions of others.³⁰ Given that today societal-level selection pressures seem, at best, weak, the organization of the complexity of the Open Society must derive largely from this “bottom-up” process—a form of self-organization.³¹

Self-organizing systems are a special subset of dynamical systems. The hallmark of self-organization is the emergence of order from the interactions among a typically large number of components without any centralized control. . . . In cases of pure self-organization there is no real centralization of information or control, but the behavior of each affects that of the others in a manner that produces an overall appearance of deliberately coordinated activity.³²

The invisible hand is, of course, the most famous model of self-organization in the social sciences.³³ These systems seem truly *self*-organized: rather than being formed by the pressure of competition with other groups, each individual acts in a way to adjust her activity to that of others, producing a functional order.

Wilson believes that such systems could only be functional by chance because of what might be called his “evolutionary mindset”: at each level, actors are seeking to maximize their individual fitness in competition with others, and so unless this competition is suppressed by a higher-level selection, there is no reason to think that

²⁸ On a simple “macro” account there is only group-level selection, but then it is axiomatic that selection must be at the group level—itself a highly controversial claim (§11.4).

²⁹ This is an implication of Wilson's view. “Two Meanings of Complex Adaptive Systems,” pp. 42–6. Small-scale societies continue to be subjected to stronger macro-selection pressures. See Richerson and Boyd, *Not by Genes Alone*, pp. 206–9.

³⁰ Wilson, “Two Meanings of Complex Adaptive Systems,” pp. 32, 38ff. Cf. Volker: “Complexity scientists can demonstrate by computer simulations that collective order may emerge purely from local interaction at the micro level, without any need of central control.” “Governance and Complexity,” p. 135.

³¹ As explained in section 11.4, this is consistent with a multilevel analysis of cultural evolution.

³² Ismael, *How Physics Makes Us Free*, p. 19. See also her “Self-Organization and Self-Governance.”

³³ Not surprisingly, macro-evolution theorists tend to be dismissive of the idea of the invisible hand unless it occurs in the context of a system with fairly strong selection at the group level. See Wilson, “Two Meanings of Complex Adaptive Systems,” p. 44; Gowdy et al., “Shaping the Evolution of Complex Societies,” p. 331. However, models that understand evolution in terms of learning rules that lead agents to adopt more advantageous cooperative strategies and norms (self-organized systems) seem well described as “the invisible hand of evolution.” See Alexander, *The Structural Evolution of Morality*, p. 18.

group functional cooperation will emerge—it would be only a random event.³⁴ Our analysis in Part I is critical at this juncture, for we have arrived at a more cooperative, more cultural, picture of humans as social beings. On this view, humans are *now*, to a significant extent, strong reciprocators: we tend to respond to cooperation with cooperation and tend to inflict punishment on those who cheat on cooperative rules. At this point in our evolutionary history, the assumption of predominant antisocial self-interestedness is no longer sensible. While, of course, we are sometimes selfish (and a few of us are almost always selfish), most are often willing to cooperate when others cooperate.

To oversimplify, we overwhelmingly search out opportunities to play Stag Hunts rather than outsmart each other in Prisoner's Dilemmas (§8.3). We should not be misled by the description of the basic Stag Hunt in which the equilibria are that we do “the same” thing—hunt stag or hunt hare. That is just an artifact of the way the strategies are described. What is crucial is that we can either enter a beneficial joint activity (“hunt stag”) or draw back into a safer, individual (or secure, traditional) activity (“hunt hare”). New niches are constantly formed because individuals are discovering new possibilities of new types of joint activities, but it can be risky to engage in them. When I occupy a new niche (say, invent a new app, or write a book on complexity and the Open Society), the opportunity arises for others to use the innovations in their own projects (the supply side); suspecting that others may welcome such a joint activity (at least for the app), I search out the adjacent possible to supply it (the demand side). The point is that all of this is premised on searching out spaces for joint activities, encouraging people to leave the relative safety of individualized or traditional activity to secure the benefits of the ever-expanding possibilities of the Open Society. To be sure, joint activities always can give rise to opportunistic defection and exploitation: norms to ensure reciprocity rather than exploitation have always been critical to human cooperation. But we are not self-interested Machiavellians looking to cheat at every opportunity. As a cooperative species, defection and cheating are not the normal default: we search out, and generally do our part in, mutually beneficial exchanges. That is how our complex society has evolved, and it is the foundation of its functionality.

In human societies, as new niches are being constructed, each is reflexively responding to the actions of others in her cooperative network, evaluating whether their interactions continue to have the requisite cooperative payoffs. For the most part, we neither seek to cheat nor do we continue with activities that no longer provide benefits. Hence there are constant adjustments in the cooperative behavior of each: when, as in the gales of creative destruction, opportunities for cooperation suddenly vanish, we commence searches for new reciprocally beneficial interactions. Functionality, understood as engaging in cooperative rather than mutually destructive interactions, is secured by ceaseless individual searching for cooperative games, and re-evaluating the games they are playing, monitoring them to determine (to put it simply) whether what was previously a Stag Hunt has degenerated into a Prisoner's Dilemma. Because so little of our welfare can be achieved without cooperative relations, we constantly

³⁴ Wilson, “Two Meanings of Complex Adaptive Systems,” p. 40.

seek them out; because we are not prone to defect on them, when we do find them, they can achieve a moderate stability. As I argued in Part I (§8.3), Smith's analysis of the market order presupposes this deep human tendency to live through trades. Perhaps the most fundamental failing of both economics and traditional political philosophy has been to commence their analyses with egoism, for under egoism effective self-organization will always seem mysterious, and markets easily seen as predatory.³⁵

17.5. Too Much Complexity?

I conclude that the idea of a self-organized society based on the search for cooperative games is by no means mysterious—macro selection is not required for social order. This answers Wilson's main challenge. But, still, the challenge emanating from Simon remains (§17.3). Macro evolutionary accounts provide a clear mechanism limiting the complexity of the system: systems that are too complex, in which a change in one element affects many others (i.e., do not possess sufficient decomposability), cannot successfully evolve in a competitive environment with other systems.³⁶ If individuals are always creating new niches that introduce yet new connections between new elements, how can the Open Society avoid Kauffman's "complexity catastrophe?" Recall that on the most basic *NK* model, as *K* approaches *N*-1 (where each element is linked to all others) the system degenerates into chaos. How, then, does the ever-increasing complexity of the Open Society avoid such chaos? It would appear that the inevitable end of ever-growing complexity is a society where *K* equals something approaching *N*-1. Here we get into rather deep waters. Let me first sketch a theoretical explanation drawn from Kauffman, and then some rather more accessible insights from Arthur.

A Coevolutionary Analysis

The most basic version of the *NK* model is usually presented as static: it is typically modeled with a fixed number of elements and connections, so that increasing complexity is understood as increasingly binding each element to the others. Obviously, if *K* is constantly increasing, there would be no way to avoid complexity catastrophe. We can make the model more adequate by employing coevolutionary analyses, where two systems are not only internally linked, but also linked to each other. Kauffman calls this an *NK(C)* and *S* model, where we have *S* number of *NK* systems, linked by *C* interconnections.³⁷ Suppose we have systems *S*₁ and *S*₂, relatively small, closed economies. We assume, as our problem dictates, that each system is characterized by high *K* (many interconnections); now we also suppose, consistent with our characterization

³⁵ Cohen, *Why Not Socialism?*, p. 83. Interestingly, this is a case where the orthodox economist's "egoistic" view of the market invites the leftist critique. One error begets another.

³⁶ In philosophy, Buchanan and Powell rely on this type of argument to show the limits of complexity. *The Evolution of Moral Progress*, pp. 263ff.

³⁷ Kauffman, *The Origins of Order*, pp. 243ff.

of the Open Society, that the two closed systems have forged a low number of connections (say they have agreed to a modest trade pact). So we have two high NK systems linked by low C connections. Now recall that on our analysis of the Open Society there is a constant process of niche creation. When a niche is created, then, it can either link to a HIGH number of other elements *in its own system* or a LOW number. Adopting an evolutionary analysis, we can say that over the long term, niche creation will tend to focus on those possible niches that have higher value for the creating agent. Now Kauffman's important work indicates that under these conditions, the niches that have the highest value in each system will be those with LOW connections. That is, in two coevolutionary systems, each characterized by high K , coupled together with low C , LOW connection niches will be more adaptive than HIGH connection.³⁸ These niches, which are less sensitive to the internal factors, are better able to adjust to changes in the coevolving other system and gain its benefits. Reducing sensitivity to domestic companies may help a firm in S_1 take advantage of activity in S_2 . Very roughly indeed, we might say that it pays a firm to reduce domestic connections to take advantage of foreign ones. So by introducing modest links between two high K systems, adaptive considerations push toward reducing the K within each system.³⁹

In this case, connecting two highly complex systems *decreases* the internal complexity of each as a new, larger, complex system develops and the subsystems adjust. Our two small economies may have been extraordinarily chaotic, as a shortage in one area fed into all others. But uniting them into a larger system, they move toward less internal complexity. So although overall system decomposability has decreased, so has subsystem complexity.⁴⁰

Now given the process of ever-increasing complexity of the Open Society, we should expect that the C connections between the two systems will themselves eventually increase. In this case, as C increases, HIGH connection niches within S_1 and S_2 will become more adaptive (they will be of high value to niche-creating agents), and so the systems will, as it were, "recomplexify" themselves. Kauffman's analysis suggests that over the very long term the systems will settle in a state of "self-organized criticality"—a state of high complexity at the edge of chaos.⁴¹ At this state the systems tend to be stable, while other systems continue to evolve. Yet such states are also subject to periodic destructive cascades as, because of either exogenous or endogenous factors, they cross the barrier from self-organization to chaos. In such cases, reorganization must recommence, and complexity again grows. But remember that all the while, the frontiers of the adjacent possible are expanding, so what was our larger coevolved system itself becomes linked into an ever-larger network, thus reintroducing the complexity-reducing dynamic in the previous paragraph. In a dynamic system there is no stopping point in this dance of lower system and increasing subsystem complexity, but the overall trajectory is an ever-larger system of interactions.

³⁸ This does not depend on a group selection mechanism. Ibid., p. 252. This analysis is a modification of Kauffman's genetic selection model.

³⁹ Ibid. This work is based on computer simulations. Notice that we are allowing K itself to evolve here.

⁴⁰ Those who believe that autarkic economic systems are more stable than a globalized economy might reflect on this.

⁴¹ Kauffman, *The Origins of Order*, pp. 255ff. See also Miller, *A Crude Look at the Whole*, chap. 11.

The Coevolution of Technology

Brian Arthur, studying technologies, also advances a coevolutionary analysis. Arthur distinguishes two facets of evolutionary complexity: that of the complexity of the structure of a technology and that of the ecosystem in which it occurs. Technologies, as we have seen, tend to become individually structurally more complex. We wish the technologies (including our social systems) to improve the range at which they operate, their robustness in relation to unanticipated events, the way they service other technologies, and their own reliability. All these tend to increase the complexity of the technology. At the same time, the diversity in the related ecosystems grows, as new niches are called forth to complement the developing technology: as we have seen, the deepening of computer technology called forth a wide array of new niches, including printers, programs, furniture, communications, changes in work routine, and so on. However, as systems develop, they sometimes confront a complexity dead-end. By the 1930s, piston aircraft engines became “incurably complicated”; they had reached the limits of speed and altitude and were unable to be successfully further developed.⁴² When they were replaced by the jet engine, there was a reduction of complexity: the ecosystem that had developed around the piston engine was wiped out in a gale of creative destruction, and the jet engine itself was initially a remarkably simpler technology, initially giving rise to a simpler and less diverse ecology. So, as in Kauffman’s analysis, complexity can increase to the extent that the system collapses. We have already seen the next stage of this tale in aircraft engines, in which the jet itself develops structural complexity within a diverse and increasingly complex ecosystem. “In this way the growth of coevolutionary diversity alternates in a sporadic way with the growth of structural depth in the strategies.”⁴³ Arthur concludes that “in evolving systems, bursts of simplicity often cut through growing complexity and establish a new basis upon which complication can again grow. In this back-and-forth dance between complexity and simplicity, complication usually gains a net edge over time.”⁴⁴ To say that the Open Society is one of ever-increasing diversity and complexity is not to say that it is on a constant gradient of increasing complexity, but that it displays an overall increase in diversity and complexity in the midst of periodic bursts of re-simplification that recommencing the growth of diversity and complexity.

§18 The Morality of Self-Organization

18.1. Functional Morality in a Self-Organized Society

Our analysis of self-organization certainly does not imply that it proceeds willy-nilly, as if any collection of individuals will successfully organize themselves into a cooperative,

⁴² Arthur, *Complexity and the Economy*, p. 174.

⁴³ *Ibid.*, p. 175.

⁴⁴ *Ibid.*

functional, social system. Some sorts of complexity—some types of interactions among individuals—manifestly impair cooperation. Again, think of our game model: if individuals can successfully seek out Stag Hunts, we can expect fruitful self-organization, but if conditions or norms somehow lock them into Prisoner's Dilemmas, we would expect dysfunctional relations rather than self-organization. Hobbes did, after all, have a point. To say that cooperative relations are fruitful and attractive to humans is not to say that they are guaranteed. To say that the Open Society is one of ever-increasing diversity and complexity is not to say that all complexity is consistent with it. We need to inquire into the conditions that facilitate the sort of bottom-up self-organization we have been analyzing. Social morality is critical in this regard.

We saw in Part I that the key of ultra-social life under conditions of disagreement is reconciliation on shared rules (§9.3). It has never been the case that humans were able to live together because they simply shared common goals; we are primates, not ants, and so cooperation always needs to be reconciled with sharp differences and conflicts.¹ Socially shared moral rules, it will be recalled, allow humans to develop both the common expectations and practices of accountability on which effective cooperation depends. The moral rules of a complex society serve to dampen its complexity by constraining some of the ways that we adjust to each other's activities, providing us with some firm expectations in the midst of constant adjustments.² As Hayek insisted, without shared moral rules the highly diverse reflexive actors of the Open Society could not even begin to effectively coordinate their actions.³ Shared moral rules allow for significant prediction of what others will do—or, more accurately, not do.⁴ Yet, at the same time, while providing expectations on which to base planning, they must also leave individuals with great latitude to adjust their actions to the constant novelty which complexity generates. These two desiderata push in opposite directions: one toward stability of expectations, the other toward freedom to change them. Successfully securing both is the main challenge of the morality of an Open Society.

All this supposes, of course, that moral norms continue to have a critical functional role in the Open Society (§13.2).⁵ Just as our evolutionary account focused on the role of morality in securing and expanding ultra-social cooperation, so our analysis of the morality of the Open Society focuses on its functions in organizing its autocatalytic complexity. This is not to say that we are to evaluate our rules purely by their functionality (§19). Like any technology, the morality of the Open Society can be evaluated on a number of dimensions. The 1957 Chevy Bel Air was a striking automobile, and its chrome and fins were aesthetically (at least to me and my brother) wonderful. Yet while its evaluation could not be reduced to the functional, the functional was critical: if it could not perform the expected role of a car in our life, then it was unacceptable, fins or no fins. The same is true of morality.

¹ "There can be no human society without conflict: such a society would be a society not of friends but of ants. Even if it were attainable, there are human values of the greatest importance which would be destroyed by its attainment, and which therefore should prevent us from attempting to bring it about." Popper, *Unended Quest*, p. 116.

² "By channeling choices into a smaller set of actions, institutions can improve the ability of an agent to control the environment." North, *Understanding the Process of Economic Change*, p. 13.

³ Hayek defined equilibrium in terms of the compatibility of plans and expectations. Hayek, "Economics and Knowledge."

⁴ *Ibid.*

⁵ See Van Schoelandt, "Functionalist Justice and Coordination."

Like any sophisticated technology, the technology of cooperation and social living we call “morality” has interconnected elements. Each of these has its distinctive role to play in a morality organizing the complexity of the Open Society. In this section, I enumerate some moral elements that are required for the Open Society. The thought is that in some form or other, some instantiation of these elements must be part of a well-functioning Open Society. In sections 19 and 20, I turn to the vexed issue of the justification of specific forms.

18.2. “Property and Right”

According to Hume, the “partition of employments” presupposes justice: there is no possibility of exchange without delimiting endowments. Justice commences when “every one has acquired a stability in his possessions, [and so] there immediately arise the ideas of justice and injustice; as also those of property, right, and obligation.”⁶ The foundations of justice are thus rights of the person and property. Unless individuals have endowments over which they have recognized claims, they cannot enter as agents into the Open Society. Unless persons has a wide scope to act as they see fit, they cannot explore the adjacent possible. We can understand Hobbes’s state of nature as a proof of the impossibility of self-organization when there is “no propriety, no dominion, no mine and thine distinct; but only that to be every man’s, that he can get, and for so long, as he can keep it.”⁷

The Open Society is the result of the self-organization of countless individuals, exercising their rights as free and equals, with the recognized normative status to use and exchange resources. Participation in reciprocal exchange and joint activities presupposes that one is an agent with rights of person and property who is authorized to decide whether to engage in, or refrain from, such activities. A normative order based on exploring and exploiting the adjacent possible must specify the rights and endowments of these explorers and exploiters. For effective self-organization to occur, the constituent agents must have this recognized agential status. Moreover, as North notes, the development of well-specified property rights helps individuals to cope with the novelty and uncertainty of a complex order, by adding control and predictability to individuals’ lives.⁸ Societies in which common property dominates, it seems, tend to be closed in the sense that they are not hospitable to newcomers.⁹

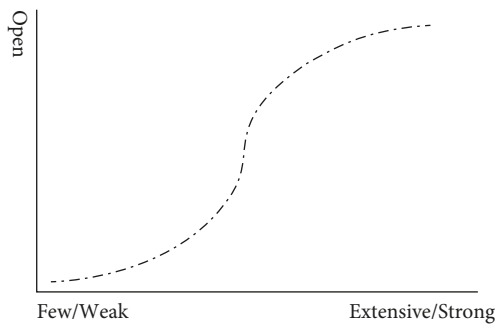
⁶ Hume, *Treatise of Human Nature*, pp. 49–91.

⁷ Hobbes, *Leviathan*, p. 78. A fundamental misunderstanding of the “Coase Theorem” is that specification of property rights does not matter, since regardless of how the rights are distributed there can be Pareto-optimal bargains as to who will incur the costs and reap the benefits of activities. On this erroneous interpretation, Pareto-optimal distribution of costs and benefits could occur in Hobbes’s state of nature. This is not Coase’s point. Coase’s argument is that all property rights determine who will bear the cost or incur the harm, and we can see with zero transaction costs how parties will distribute the harms of an activity. In world of significant transaction costs, property rights matter a great deal, and ideally should be the best estimate of how, socially, the gains and losses of an activity should be distributed. Also, note how “bargaining” for Coase does not include threats and force; thus it too supposes a background of norms. See Coase, “The Problem of Social Cost.”

⁸ North, *Understanding the Process of Economic Change*, p. 15. At the same time, the uncertainty of other people can be increased because they are subject to the decisions of property owners.

⁹ Platteau, *Institutions, Social Norms, and Economic Development*, p. 82.

That the Open Society cannot exist without extensive individual property rights that clearly identify endowments and freedom to employ them does not imply that all resources, much less all decision-making, is to be privatized. Whether privatization is the best way to make resource decisions depends on the nature of the resources and the moral norms of a population.¹⁰ Consequently, it is mistaken to claim that the Open Society requires “full liberal ownership rights” over the maximally large range of resources.¹¹ This merits emphasis: that robust individual property rights are required for the Open Society does not entail that expanding the sphere of private property is always friendly to the Open Society. We might hypothesize an S-shaped function, as in Display II.2.



Display II.2. Hypothesized relation between strength of property rights and the Open Society.

When, as in parts of sub-Saharan Africa today, physical and social conditions frustrate the development of individualized property rights such that they are few and weak, the rise of markets and entrepreneurial activity is almost entirely thwarted.¹² Historical studies indicate that it is not until extensive development of individualized private property that effective and widespread markets arose—incomplete property rights specification seemed critical to the failure of Russia’s attempt to institute a market economy.¹³

The resistance to the development of property rights in sub-Saharan tribal societies points to one way in which, indeed, the morality of the Open Society is inconsistent with the sort of strict egalitarianism characteristic of many forager and

¹⁰ See Ostrom, *Governing the Commons*; Platteau, *Institutions, Social Norms, and Economic Development*, chaps. 3 and 4.

¹¹ These typically are held to include the right to use, the right to exclude others, the right to manage, the right to compensation, the rights to destroy, waste, or modify, the right to income, immunity from expropriation, liability to execution, absence of term, to rent and sale. For an argument against insisting on full liberal ownership, see my essay on “Property.” For an analysis of “maximal capitalism,” see my “The Idea and Ideal of Capitalism.”

¹² For a careful study, see Platteau, *Institutions, Social Norms, and Economic Development*, chaps. 3 and 4.

¹³ See North, *Understanding the Process of Economic Change*, chap. 10–1.

small-scale societies (§7.3). About this Hayek was undoubtedly correct (§2). Recall that these societies enforce strict equality through a type of counter-domination: individuals are rebuked for any indication that they can lord it over others and dominate them. As Platteau describes many sub-Saharan tribal societies, they are far more concerned with relative equality than increasing absolute gains, and employ coercion to protect it. These societies are both fairly repressive (there is strict social monitoring of other's resources), and employ "forced generosity" when some do accumulate. Entrepreneurship and exploration are thus strongly discouraged, being seen as treacherous exploitation.¹⁴ And, partly as a result, these societies remain appallingly poor. According to Abraham and Platteau, to alleviate this poverty, "what is required is a shift from a status-based and coercive society that relies on mutual control, respect of ranks, and strictly enforced codes of generosity, to an open society where free entry and exit, democratic governance (including acceptance of dissent), competence criteria, and socioeconomic differentiation are used as guiding principles or expressly allowed to operate."¹⁵

Egalitarian norms, however, need not take the extreme leveling form of many sub-Saharan tribal societies. As Platteau argues in the development context, "egalitarian norms in Asia are not loaded with egalitarian beliefs as strong as in Sub-Saharan Africa. Rather, they are based on the claim 'that all should have a place, a living, not that all should be equal.'"¹⁶ Note how this echoes our evolutionary analysis, which stressed the fundamental role of reciprocity in fair relations in the development of morality, and the possibility for reconciliation of liberty and equality. Recall in this regard the social insurance scheme of the Ache (§8.3).

As I have stressed, it is a serious error to assume that property rights and markets are grounded in egoism. They not only depend on norms of honesty and trustworthiness, but as we have seen (§13.2), the well-known studies of Henrich and his co-investigators indicate that markets are associated with norms of fairness to strangers.¹⁷ Whereas small-scale non-market societies play Ultimatum Games in diverse ways (including relentlessly "egoistic" ways), all societies acquainted with markets display comparable reciprocity-based fairness behavior by Proposers. As Platteau observes, this is connected to the impartiality of market morality:

Max Weber, as is well-known, has contrasted the achievement of generalized morality (what he called "universally binding morality") in western Europe in modern times with the limited domain of trust in traditional societies. Weber's point is indeed that the universal diffusion of unscrupulousness in the pursuit of self-interest was far more common in precapitalist societies than in their more competitive capitalist counterparts.... While normative behaviour based on the principle of reciprocity ("I cannot expect others to deal honestly with me unless I am honest with them") was restricted to real kith and kin in the former societies, security of expectation or

¹⁴ Platteau, *Institutions, Social Norms, and Economic Development*, pp. 206ff.

¹⁵ Abraham and Platteau, "Participatory Development in the Presence of Endogenous Community Imperfections," p. 26. Compare North, *Understanding the Process of Economic Change*, pp. 100ff.

¹⁶ Platteau, *Institutions, Social Norms and Economic Development*, p. 235, quoting James Scott. Compare North, *Understanding the Process of Economic Change*, p. 58.

¹⁷ See also North, *Understanding the Process of Economic Change*, p. 46.

assurance has been provided by the pervasive influence of a code of generalized morality in the latter.¹⁸

It would be absurd to identify a single or even a small set of factors that drove this expansion of the impersonal morality of generalized reciprocity, but our strategic model from section 10.3 highlights a critical dynamic. As groups began to expand for a variety of internal and external reasons, conflict between competing moralities was bound to arise. Thus the “taking sides” aspect of morality rose to the fore, and those moral views that could “climb the ladder” of impartiality were better able to serve as the morality of the larger group. A sustained pressure for increasingly impartial moralities favored the development of the Open Society’s expansion of the basic rights of persons and property, providing a structure for the ever-widening scope of moral cooperation. Thus, for the liberal morality of the Open Society, increasing impartiality and moral inclusivity become the criterion of moral progress. “Among the most important moral beliefs, from the standpoint of moral inclusivity, are those concerning moral standing and equal basic moral status . . . slavery and other forms of bondage, as well as discrimination on grounds of gender, ethnicity, or religion, are morally wrong.”¹⁹

Once again we arrive at a complex conclusion. On the one hand, Hayek is correct that small-scale bands and tribes display features fundamentally at odds with the Open Society. They often rely on a harsh reverse-dominance hierarchy to impose a strict egalitarianism, and operate on largely personal rather than impersonal and impartial norms.²⁰ But that is but one form of evolved human egalitarian relations. Norms of reciprocity and cooperation have allowed humans to quickly increase the sphere of cooperation far beyond the band from about 15,000 years ago (§12). The fundamental innovation of the morality of the Open Society is its tremendous scaling up of the norms of reciprocity and fairness, while incorporating the ancient concern with autonomy and personal freedom (§7.4): the core rights of the person and property become universal. To be sure, inequality of property and its consequences are always in tension with this egalitarianism: the rich often dominate and bully. But we have seen that human morality always has displayed antinomies. Perhaps that, as much as any feature, defines human morality.

18.3. The Moral Rules of Interaction: A System of Natural Liberty

On certain “rights-based” theories, once we specify the endowment rights of individuals, the normative terms of social interaction are set. In Nozick’s theory, for example, it seems that once the set of individual rights is specified, we have the full set of moral claims that one person has against another. He explicitly voices skepticism

¹⁸ Platteau, *Institutions, Social Norms and Economic Development*, p. 302, citation omitted.

¹⁹ Buchanan and Powell, *The Evolution of Moral Progress*, p. 12.

²⁰ These are by no means simply Hayekian claims: both are stressed by North, *Understanding the Process of Economic Change*, and Platteau, *Institutions, Social Norms and Economic Development*, chap. 5.

regarding whether there is room left for further normative evaluation of social relations once all individual and property rights have been specified.²¹ Now, to be sure, there is no clear line between one's core rights of agency and property and the myriad social rules one lives under: what constitutes one's property is affected by rules against trespass, nuisance, offenses, and so on. Yet to conflate all moral norms with rights of person or property obscures the critical fact that in the Open Society agents are constantly negotiating the terms of their interactions without negotiating away their core standing and endowments. What constitutes an externality, an insult, a failure to tip, irresponsibility as a professor, culpable ignorance, or bullying? What normative expectations should parents have of children (and vice versa), of spouses toward each other, of shareholders and stakeholders to each other and to the public? To reduce all these issues of normative regulation to the shape of one's agency and property rights is to purchase theoretical elegance at the price of practical confusion.

On Smith's system of "natural liberty," all rules of social morality are rules of justice, protecting the property and person of individuals, as well as rules concerning their reputations. He and Hume are, as we have seen, certainly correct that the basic rights of justice are fundamental to a society based on the division of labor, but they do not exhaust the morality of the Open Society. But Smith's idea of a system of natural liberty is insightful in another way. For Smith,

Mere justice is, upon most occasions, but a negative virtue, and only hinders us from hurting our neighbour. The man who barely abstains from violating either the person, or the estate, or the reputation of his neighbours, has surely very little positive merit. He fulfils, however, all the rules of what is peculiarly called justice, and does every thing which his equals can with propriety force him to do, or which they can punish him for not doing. We may often fulfil all the rules of justice by sitting still and doing nothing.²²

As Shaun Nichols and I have argued, negative, or prohibitory, rules are especially appropriate to the Open Society.²³ Prohibitory rules seem especially adept as facilitating the constant searching for, and learning about, new niches that are constitutive of the autocatalytic diversity of the Open Society. A prohibitory rule such as "In circumstance C,²⁴ Do not litter!" helps trim the set of eligible options open to a complying agent, but by no means determines action.²⁵ There are innumerable ways of not littering (including, as Smith pointed out, by sitting still and doing nothing.) The benefit of this is that such rules, while giving others firm expectations about what you will

²¹ Nozick, *Anarchy, State and Utopia*, p. 166.

²² Smith, *The Theory of Moral Sentiments*, p. 82.

²³ Gaus and Nichols, "Moral Learning in the Open Society." See also my *The Tyranny of the Ideal*, pp. 187–98.

²⁴ To know a rule is to know the circumstance that triggers it. See Cialdini, Kallgren, and Reno, "A Focus Theory of Normative Conduct"; Bicchieri, *The Grammar of Society*, chap. 2; Bicchieri and McNally, "Schemata, Scripts, and Social Norms"; Gaus, *The Order of Public Reason*, pp. 167ff.

²⁵ Talk about "trimming" the set of eligible options is useful here, but given our flexible moral consciences, the options are never really removed, i.e., "trimmed." Rather, they become radically discounted. For a fuller and more nuanced analysis, see *ibid.*, pp. 131–60.

not do, nevertheless allow you to explore new possibilities.²⁶ Contrast this to a permission rule, according to which one “may ϕ .” The very nature of the rule presupposes defined action descriptions: one can only give a permission to do what can now be described.²⁷ In this sense, permission rules seem most appropriate to relatively stable social orders, where the set of available action options can be fairly well identified.²⁸

A second feature of prohibitory rules entails Smith’s notion of “Natural Liberty.” Systems primarily composed of prohibitory rules are “closed” by an unspoken rule according to which if there is no rule prohibiting an action, it is permissible to perform it.²⁹ A human rule system in a complex society requires such “closure” rules if it is to be decisive—if it is to always provide guidance.³⁰ For a rule system to be decisive, it must be complete in the sense that for any act ϕ in circumstance C , the rule system determines whether the act is permitted or prohibited.³¹ When a moral rule system is decisive, it never declares that social morality simply has nothing to say as to whether one may (or may not) perform an action: a non-decisive moral system declares that there *is* no correct moral answer to whether one may ϕ . Moral rule systems that are incomplete in this way are seriously flawed as frameworks regulating social cooperation as they provide no moral guidance at all for actions not specified in the rule set, and so entirely fail to coordinate expectations. And given that a practice of accountability is part and parcel of social morality (§9.4), non-decisive rule systems encourage social conflict as to whether a person is accountable for an action. A closure rule guarantees decisiveness: what is not spoken to by the rule system is put into a residual category, which is then either permitted or prohibited. In the case of Smith’s system of Natural Liberty, *what is not prohibited by a rule of justice is permitted*. “Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest his own way.”³² Smith’s system of Natural Liberty is psychologically well-grounded: when faced with a system of prohibitions—or, importantly, a system of prohibitions mixed with permissions—ordinary reasoners have a strong tendency to apply the Natural Liberty closure rule.³³ They see the need for a closure rule, and strongly tend to apply Natural Liberty. In contrast, when people are taught a system of *solely permissive* rules,

²⁶ “One of the main aims of the rules . . . must therefore be to eliminate for the individual as much avoidable uncertainty as possible. This means that he must be able to ascertain from the circumstances which he can know what he is free to do, and under what circumstances and in what manner other human forces will constrain him. If he is to use his knowledge to the best advantage for achieving his aims, the world around him must be, as far as possible, given to him. Of course, in a changing world much of his task and his merit will be to foresee changes correctly, to adapt himself successfully to ever changing conditions.” Hayek, “The Political Idea of the Rule of Law,” p. 162.

²⁷ Some permissive rule systems can be open to extensive action options. Chess is defined by a set of permissive rules, defining what moves are permitted by each piece. The number of possible moves in chess has been estimated as at least 10^{120} . In light of Kauffman’s analysis (§16.4), compare a rule system that permitted certain uses of a screwdriver (to screw screws, to chip paint) to one that prohibits some uses (in murder).

²⁸ In Russell Hardin’s term, permission rules seem appropriate to the “Bodo ethics” of small traditional societies. See my “The Priority of Social Morality.”

²⁹ Nichols and Gaus, “Unspoken Rules”; Mikhail, *The Elements of Moral Cognition*.

³⁰ The two implied caveats: “human” because we can imagine demi-gods who might be able to enumerate the entire set of rules covering all cases, “complex” because in very simple society perhaps humans could do this. See Gaus and Nichols, “Moral Learning in the Open Society,” for consideration of these issues.

³¹ “ ϕ is required” can be analyzed as “not- ϕ is prohibited.”

³² Smith, *Wealth of Nations*, vol. II, p. 208.

³³ Nichols and Gaus, “Unspoken Rules”; Gaus and Nichols, “Moral Learning in the Open Society.”

they strongly tend to infer a “residual prohibition principle”: that which is not permitted is prohibited. The conservative nature of such a closure rule is manifest.³⁴

Natural Liberty rule systems are thus inherently open-ended: what is not spoken to is permitted. In a social world where the possibilities for action are constantly expanding, we cannot comprehensively say what actions morality permits.³⁵ Natural liberty is fundamental to the moral rules of an Open Society, in which forays into the adjacent possible are seldom predictable. As Hayek observed, “Freedom granted only when it is known beforehand that its effects will be beneficial is not freedom. If we knew how freedom would be used, the need for it would largely disappear.”³⁶ In contrast, moral rule systems that rely on the residual prohibition stifle autocatalytic diversity: if successful, they eliminate the exploration of the unknown in the interests of predictability and control. As an editorial of *The Economist* pointed out in relation to the liberalization of Cuba, “In place of a ‘positive list’ of permitted private activities, the government should publish a negative one that reserves just a few for the state. All others would then be open to private initiative, including professions such as architecture, medicine, education and the law.”³⁷ This, however, would unleash the forces of autocatalytic diversity, something the Cuban government adamantly opposes. The enemies of Natural Liberty are, in the end, the enemies of the Open Society.

The core claims of this argument for Natural Liberty, then, are five:

- (i) Humans require shared social rules to coordinate their expectations to secure cooperation.
- (ii) We cannot learn a set of rules that cover all possible situations. Not only would such a set far exceed our cognitive limitations of learning and teaching rules, but enumerating all possible situations is itself an incoherent endeavor in a society characterized by autocatalytic diversity.
Consequently,
- (iii) humans require rule systems with closure principles, which cover cases not spoken to by the rules, and
- (iv) only the Natural Liberty closure rule is consistent with the search of the ever-expanding adjacent possible.
Additionally,
- (v) as Smith indicated, the closure principle of Natural Liberty is best understood as part of—and is associated by normal reasoners with—a rule system focused on prohibitions, hence the System of Natural Liberty.

18.4. The Moral Rules of Interaction: Harm Rules

The dynamic of moral expansion sketched in section 18.2 also helps explain why the morality of the Open Society is basically a morality focused on fairness, harm, and

³⁴ See Gaus and Nichols, “Moral Learning in the Open Society.”

³⁵ See further Rose, “On the Evolution of Ethics, Rationality, and Economic Behavior.”

³⁶ Hayek, *The Constitution of Liberty*, p. 31.

³⁷ “Reforming Cuba,” *The Economist*, May 16, 2015.

freedom—that is, the morality of Western Educated Industrialized Rich Democracies (§12.4). As morality becomes increasingly impartial, many of the richer features of morality that depend on ethnic traditions, kin networks, and religion fail to climb the ladder of impartiality. Subgroups continue to uphold them in their narrower forms of shared life, but they are ineffective in gaining the adherence of wider groups in competition with those moralities that focus simply on the core features of agency and cooperation. We have seen that these core cooperative concerns are as old as human morality itself. Thus, rather than seeing so-called WEIRD morality as a Western invention, it is more aptly seen as a thinning out of social morality to include only the critical features necessary to functional human cooperation.³⁸

Having already considered liberty and property rules, it is appropriate here to stress the importance of harm rules (for fairness, see §§18.5, 21.3). In our experiments on people's use of the Natural Liberty closure rule, Nichols and I found that the strong supposition that what was not prohibited is permitted was largely overturned in contexts where the action involved harm to others.³⁹ Subjects were taught on a system of prohibitions that did not mention cases of harm, and then were asked to judge the permissibility of an action that involved harm. On the Natural Liberty closure rule, what is not specifically prohibited is permitted. Yet we found that actions involving harm are viewed as impermissible, even if the rule system does not prohibit them. Harm to others effectively checks the invocation of Natural Liberty: even when told that the rules are complete, subjects apparently invoked an anti-harm principle.

Recall that the aim of *On Liberty*

is to assert one very simple principle, as entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control, whether the means used be physical force in the form of legal penalties, or the moral coercion of public opinion. That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant.⁴⁰

While our study confirms the importance of a harm principle in normal subjects' judgments of permissibility (overriding their commitment to Natural Liberty), it does not of course support the claim that it is the "sole" principle of moral regulation. More importantly, although subjects tended to think harm to self was somewhat more permissible than harm to others, this difference was not significant. *Pace* Mill and much of the liberal tradition, there was a strong tendency to see harm to self as also something to be prohibited, even when the rules do not prohibit it. Here, perhaps, we find a significant gap between folk intuitions and the liberal program. Interestingly,

³⁸ "The resistance against the new morals of the Open Society was strengthened also by the realization that it not only indefinitely enlarged the circle of other people in relation to whom one had to obey moral rules, but that this extension of the scope of the moral code necessarily brought with itself a reduction in content." Hayek, *The Mirage of Social Justice*, p. 146.

³⁹ Nichols and Gaus, "Unspoken Rules."

⁴⁰ Mill, *On Liberty*, p. 223.

however, subjects were significantly less inclined to invoke a “precautionary” principle that prohibited actions with a risk of harm (that had not been shown to be harmless).⁴¹ When the potential for harm to self is explicitly said to be left unexamined, people are more likely to judge the act permissible than in harm to others, yet they nevertheless have more tendency to see it as impermissible than in cases where no possibility of harm is mentioned.

18.5. The Complex Society Moral Package

We are now in a position to give our complete answer to Hayek’s first unsettling thesis—that our evolved morality, developed in small-scale society, is fundamentally hostile to that of the morality of the Open Society. Part I explored the evolution of what I have called the Modern Egalitarian Moral Package, which includes sustained subordinate rebellion—protection of autonomy via control of dominators through a reverse dominance hierarchy, conditional cooperative motivations, a deep concern with reciprocity and fairness, ability to learn and internalize cooperative rules that treat members as free and equal, sensitivity to social rules, a flexible conscience, reconciliation on common rules, and readiness to punish offenders. And, critically, it includes a practice of accountability that requires argumentative skills and the practice of moral justification, leading to disagreements, as well as a competition to “climb the ladder of impartiality” as a way to resolve them. In this Part we have been assembling the elements of what I shall now call the Open Society Moral Package. The obvious question seems to be: Do these manifest conflicts?

But this is a confused question, for the evolved Modern Egalitarian Moral Package is itself full of tensions. The great myth is that we have a uniform and harmonious moral nature, one which is ill-suited to the Open Society—we have no such nature. The apex of human egalitarianism occurred in small-scale hunter-gatherer bands (§12.3) and today is best represented in small-scale tribal societies, such as those in sub-Saharan Africa. In one way, these egalitarian societies strongly protect the freedom and autonomy of individuals from domination by bullies and those who seek to be preeminent. They certainly have a higher degree of freedom and autonomy than the archaic despotic societies (§12). But the very way they protect freedom and autonomy against domination by the powerful tends to be highly coercive—as Platteau observes from his own development studies, such egalitarianism involves powerful sanctions against “progressive individuals.”⁴² Moreover, Hayek is certainly correct that small-group morality has barely begun to climb the ladder of impartiality, being based on personal and kin-relations.⁴³ On the other hand, different manifestations of egalitarianism, such as the reciprocity-based versions observed in Asian villages (§18.2) and among the Ache (§8.3) seem far more friendly to entrepreneurial freedom and economic growth.

⁴¹ Whereas judgments of permissibility in harm-to-self cases are not significantly different from those in harm-to-others cases, judgments of permissibility of potential harms are.

⁴² Platteau, *Institutions, Social Norms and Economic Development*, p. 189.

⁴³ *Ibid.*, pp. 208ff., 246ff.

It is also important to recognize that the Modern Egalitarian Moral Package has always existed in tension with—indeed, as a counterweight to—important features of human existence. We are self-interested and many seek status: egalitarian morality has served as counterweight to these tendencies, while also allowing many expressions of them. Overall, I think what is clear is that the Open Society Moral Package draws on the core elements of the Modern Egalitarian Moral Package, and by no means turns its back on our evolved moral outlook—protection of autonomy, conditionally cooperative motivations, a concern with reciprocity and fairness, ability to learn and internalize cooperative rules that treat members as free and equal, sensitivity to social rules, a flexible conscience, reconciliation on common rules, and readiness to punish offenders. To be sure, the complex life of the Open Society requires inequalities that sometimes generate subordination-producing hierarchies. Clear endowments and the ability to explore new possibilities through “market tested innovation” (§14.1) are at the heart of the morality of the Open Society, and these will inevitably lead to inequalities in valued outcomes for different agents. And great inequalities, especially conjoined with other moral norms and social dynamics, may pose a threat to autonomy and non-subordination. Such conflicts, and the need to manage them, have always been at the heart of human morality. Many long for a state of human existence where all is harmony: to the libertarian, we can have freedom with no worrying tendencies to inequality, bullying, and subordination; the egalitarian responds that if we only get the institutions and motivations right, we may have equality with no sacrifice of freedom and innovation, and no significant impairment of the ability of individuals to adjust their actions to the plans of others. The dynamic moral and political life of an Open Society that we shall be exploring in the rest of this Part can be understood as a response to the constant, and constantly changing, moral problems confronting the Open Society. It would be ludicrous to analyze a constantly changing society by proposing a fixed set of rules or institutions that ensure that it will not experience conflicts and will be devoid of internal tensions.

§19 The Social Contract

19.1. Impartiality

We now, finally, turn to the crux of the challenge advanced by Hayek’s second unsettling thesis. Can the rules of the Open Society be justified? If not, the Open Society Moral Package would be manifestly at odds with the Modern Egalitarian Moral Package, since justificatory arguments are so central to evolved human morality. However, given that a defining feature of the morality of the Open Society is its high degree of impartiality, and impartiality seems critical in justifications resolving moral disputes (§10.2), it would be surprising indeed if its impartiality were divorced from justification.

An especially surprising aspect of Hayek's criticism of the justificatory enterprise was his rejection of the social contract as a constructivist planning device. As noted in the Prolegomenon, of all the traditional methods of moral and social philosophy, the social contract has been perhaps most friendly to diversity. Rawls resurrected social contract theory explicitly as a way to model disagreement (about the good) within social and political philosophy. I begin by reconsidering in what ways the social contract device can assist us in thinking about the justification of the Open Society—and how it fails, using Rawls's contract as the focus.

We might distinguish two different justificatory aims of a social contract theory. Thinking about Rawls's theory is especially helpful in this regard because its two principles each exemplify one of these aims. His argument for the first principle of justice—which seeks to enumerate basic civil and political rights—is best understood as maintaining that a certain interpretation of impartiality among diverse individuals provides a normative case for a favored interpretation of the basic rights of the person.¹ I have argued that a critical feature of the morality of the Open Society has been its climb up the ladder of impartiality. Consequently, providing a normative case for some interpretation of the core rights is closely tied to showing their impartiality—what is more impartial seems able to attain the widest possible endorsement in a diverse society. Rawls's preferred interpretation derives impartiality from choice under ignorance, the “veil of ignorance.”² If one must choose principles of justice without knowing one's conception of the good (or justice, or one's own interests), then obviously one cannot be partial to them in the choice. Hence one's information is restricted to matters that all share, ensuring that the favored interpretation of the basic rights track these—and only these.

This was an important and enlightening innovation in thinking about the basic rights in a diverse society. Its weakness, however, becomes apparent if we think of “climbing the ladder of impartiality” not in terms of what is stripped of all semblance of partiality, but what can be endorsed by the widest possible public because it is not sectarian. The latter, it will be recalled, was the core idea driving our evolutionary model of impartiality: we are driven to overcome our more sectarian moralities because the benefits of the more impartial morality exceeds its costs (§10.3). It does not follow that “*P* is stripped of all partiality” implies that “*P* is the subject to the widest possible endorsement” because moral endorsement of any principle *P* is ultimately a matter of whether the moral benefits of agreement exceed the moral costs. The parties behind Rawls's veil of ignorance do not know what they must give up, or even the full moral importance of what they would obtain (since they have no knowledge of their unique moral perspective). It is thus by no means surprising that many readers reject Rawls's claim that the preferred interpretation for the fundamental rights excludes property (except for personal possessions), for on their moral perspectives this is a fundamental right that they must abandon to embrace Rawls's impartial morality.³

¹ The insight that the Rawlsian-inspired program is premised on a notion of impartiality is thoroughly explored by Barry, *Justice as Impartiality*. Barry is especially clear that such impartiality is a sort of second-order impartiality, seeking to inquire how those who disagree can impartially live together. See especially chapters 1 and 4 of *Justice as Impartiality*.

² Rawls, *A Theory of Justice*, pp. 15–19.

³ See Tomasi, *Free Market Fairness*.

Not surprisingly, an upshot of this method has been complicated and continuing debates about the stability of the Rawlsian principles.⁴ If the principles are selected in ignorance of the costs to one's moral perspective, why should we expect people to conform to them? On the analysis we have been developing, impartiality is not an end in itself: it is one of the devices that a morality employs to resolve disputes and expand the network of moral and cooperative relations in a way endorsed by diverse perspectives. Impartiality is the mark of a functional morality in a diverse society, but we must not make a fetish of it. What is important is whether, for any given a set of competing moral perspectives, each perspective ranks participating in the (more) impartial morality as better than retreating to a smaller, parochial network. Like all rational decisions, it is a matter of (moral) costs and (moral) benefits. That the Rawlsian contract abstracts from these costs renders it unhelpful in thinking how a society of diverse perspectives might come to endorse a common impartial morality.

19.2. Algorithmic Justice

The second aspect of the Rawlsian contract is more straightforwardly objectionable in a complex society. Rawlsian contractors adopt two criteria of politico-economic institutional justice: they must respect fair equality of opportunity and maximize the life prospects of the least advantaged. Now Rawls is crystal clear that he does not intend these criteria to be used as a basis for constant interventions (that is, at each time slice government adjusts opportunities and holdings to conform to them).⁵ Rather, the idea is to set up institutions that secure these ends over the long run without repeated corrective interventions.

This, then, is a procedural notion of social justice. Rawls famously distinguished two sorts of procedural justice in which the justice of the procedure derives from the justice of the outcome.⁶ In perfect procedural justice we have (i) a criterion of the just outcome and (ii) a procedure that if followed is guaranteed to produce it. Thus for the classic cake-cutting game: if one holds that an equal division is just and that people are selfish, then the procedure according to which the person who cuts the cake gets the last piece will guarantee this result. Not all procedures can perfectly track the criterion of justice of outcomes: thus it is said that jury trials *tend* to produce the just outcome. Both of these might be understood as algorithms for justice. The problem is how to produce a certain outcome (equality, people are convicted if and only if they are guilty), and then an algorithm is found that leads either precisely to that result or that result given some error tolerance.

I assume here that the outcomes *O* of interest are those familiar in the social justice literature: that income is distributed according to a pattern, that some economic class ends up maximizing its long-term prospects, that overall social welfare is maximized.

⁴ This was a core problem since *A Theory of Justice* (Part III), and came to dominate the subsequent political liberalism project. See further §21.3 in this volume.

⁵ Rawls, *A Theory of Justice*, pp. 57–8, 73–7. But see the responsibilities of the “distribution branch” of government, pp. 245–9.

⁶ *Ibid.*, pp. 73–5. Pure procedural justice, in which the only criterion of justice is that the procedure is followed, does not qualify as algorithmic justice.

It is important to stress that the concern here is relatively specific distributive outcomes—a certain socioeconomic class will always benefit the most, all will receive their just deserts, distributions will compensate for all bad luck, and so on. Hence two crucial caveats: (i) I am not considering these as possible social goals: the question is whether these goals can be secured *via an algorithmic procedure*—that these outcomes are to be reached via the members of society following some algorithm such that by following its rules society will reliably achieve O_x . We are not concerned with constant interventions at each time slice to induce O_x , as might be the case of simple pursuit of O_x as a social goal. Given the nature of the outcome, these may or may not be effective. Our interest, however, is the attempt to devise an algorithm—a system of rules, regulations, or institutions—that, if followed by some threshold number of the population of the Open Society, will produce the sort of outcome that traditional theories of social justice have advocated. (ii) Again, it is important to stress that these have typically been rather specific types of outcomes—we are not concerned with whether general predictions can be made about the general consequences of rules, such that some engender cooperation while others induce conflict. The concern is whether justice in a diverse society can identify an *algorithmic procedure to secure the relatively specific distributional results* that have characterized so much political philosophy in the last fifty years.

It is most difficult to see how, in the complex system of the Open Society, algorithms for producing such outcomes could be possible. As we have seen (§§16.4–5), the possibility space is not defined, and is expanding in an unpredictable way at an ever-increasing rate. Because they are characterized by constant reflexivity (people reacting to the outcomes that are produced by their reactions, which are then reacted to by yet others, and then others, and so on) any algorithm that produces outcome O_x is immediately confronted with the problem that agents react to O_x , upsetting O_x . A system composed of only permission rules would greatly mitigate this problem since such a system defines a possibility space of action: if an action is not in the possibility space of what is permitted, then it is prohibited. So in a permissive system the space in which the algorithm is seeking to achieve outcomes is at least bounded. Of course, this boundedness is secured by thwarting exploration of the adjacent possible, and so of the Open Society itself.

This, of course, is a version of Nozick's claim that liberty upsets patterns.⁷ Take any algorithmic procedure A , which is a set of instructions (rules) for achieving O_x . Suppose the desired outcome is achieved, but individuals adjust their actions (collectively, say the pattern of adjustment is α) to produce a new outcome that is inconsistent with O_x . Now if A could predict the α pattern of actions, it could include these as "feedbacks," altering the algorithm, seeking to bring society back to O_x . But that too will be subject to reflexive feedback; once we have a few such levels of feedback, the outcomes of the algorithms become incalculable, even if we had the information required to build them (which we do not).

It is important to stress that this result need not require violating the rules of the justice algorithm. Suppose that everyone follows the rules of the algorithm and the

⁷ Nozick, *Anarchy, State and Utopia*, pp. 155ff.

desired outcome is indeed produced—let us call this “approximate equality of incomes in a reasonably prosperous society.” Some observe the outcome and reflect on their actions and opportunity costs over the period when they were following the algorithmic rules, and as a result decide to change some of their behavior. After all, they have conceptions of the good, devotions to various causes, kin, and so on to think about. They might, for example, decide to take more time in leisure and less in work; close up businesses and become employees; save less; switch savings from stocks to bonds; leave high stress occupations; drink more (in their increased leisure); emigrate; have fewer children; listen to talk radio in their increased leisure and adopt conspiracy theories. All of these may undermine the “prosperous approximate equality” outcome, yet no one cheated on the rules.

This has been implicitly recognized by those who insist that producing such a just outcome requires people to be devoted to producing it by adopting a uniform “ethos.”⁸ The dream⁹ is that, if individuals have a thoroughly uniform view about the aims of justice and are motivated by it (an “ethos”), as well as sharing a uniform high ranking of its importance vis à vis other goods, reflexivity will be modest and, if it exists, essentially uniform. “Social rules of regulation” (algorithms) to produce the outcome would then be possible. What is important is the recognition that reflexivity can only be controlled by greatly simplifying and homogenizing human society (an ambition of many political philosophers since Plato). Such egalitarian justice is not simply a rejection of liberalism, but of a diverse and complex society.¹⁰

Still, many thoughtful people insist that some such algorithms are possible and believe that we can indeed (with powerful social science) devise procedurally justified rules of this sort. I have expressed deep skepticism that this is possible, given the very nature of a dynamic, Open Society. However, I do not wish to claim that it is unreasonable to believe that such rules are possible; indeed, I shall explicitly allow that some citizens in an Open Society may hold them (§21.3). This too is a matter about which we disagree. My claim here is that a social contract based on them—a *social agreement* as to what they should be—is unacceptable since there is a strong case that such rules are inconsistent with the nature of the Open Society. If a social contract theory includes an algorithm for achieving certain patterns of valued outcome, it cannot possibly be an object of any sort of impartial consensus.

Nothing I have said here entails that it is somehow incompatible with the nature of the Open Society that endowments (holdings) can be modified. A moral principle that assured a minimum income is not in any way opposed to the very nature of the Open Society. “A system which relies on the spontaneous ordering forces of the market, once it has reached a certain level of wealth, is . . . by no means incompatible with government providing, outside of the market, some security against severe

⁸ Famously, Cohen, *Rescuing Justice and Equality*, esp. chap. 5. See also Carens, *Equality, Moral Incentives, and the Market*.

⁹ I say “dream,” since unless they also uniformly interpret the ethos, the problem reappears. It seems hard to imagine that any human society has ever shared such a homogenous ethos, and that most definitely includes the most egalitarian societies in human history, hunter-gatherer societies. See my “The Commonwealth of Bees: On the Impossibility of Justice-through-Ethos.”

¹⁰ As Popper indicated, such Platonism is the enemy of the Open Society.

deprivations.”¹¹ Whether such minimum should be high or low is a different matter—like so much, this is a matter of moral disagreement. The Open Society depends on individuals having clear endowments on which they can act, plan, and explore: moral principles that call for ensuring these endowments meet certain levels (and whether this enhancement is provided in cash or kind) are entirely consistent with this. Note that no algorithm is involved, since there would be constant “interventions” to supplement (or tax) incomes.

19.3. The Deliberative Model and Its Shortcomings

The Order of Public Reason presented a “Deliberative Model” that sought to overcome the difficulties with a standard social contract view, in particular its information-deprived analysis of impartiality (§19.1). This Deliberative Model invoked social choice theory, and largely modeled impartial acceptability in terms of an acceptable aggregation procedure.¹² The procedure was essentially what we might call “Pareto or Indifference” rule. If for everyone, moral rule *a* is better than moral rule *b* (*a* > *b*) then from the social perspective *a* is better than *b*—it is a better shared morality. If it is not the case that all prefer *a* to *b* (or *b* to *a*), then from the social perspective the society is “indifferent” (~) between the two. In addition, an option of “no moral rule on this matter” (say, *z*) was inserted into everyone’s ordering. Thus we might get something like Display II.3.

Alf	Betty	Charlie	Doris
<i>a</i>	<i>b</i>	<i>a</i>	<i>d</i>
<i>b</i>	<i>c</i>	<i>b</i>	<i>a</i>
<i>c</i>	<i>a</i>	<i>c</i>	<i>b</i>
<i>d</i>	<i>z</i>	<i>z</i>	<i>c</i>
<i>z</i>	<i>e</i>	<i>d</i>	<i>z</i>
<i>e</i>	<i>d</i>	<i>e</i>	<i>e</i>

Display II.3. An example of an optimal eligible set.

We begin by excluding *e* and *d*, since some “Members of the Public” judge them worse than no rule at all (*z*), and so would not adopt them as a common morality

¹¹ Hayek, *The Mirage of Social Justice*, p. 136. “The assurance of a certain minimum income for everyone, or a sort of floor below which nobody need fall even when he is unable to provide for himself appears not only to be a wholly legitimate protection against a risk common to all, but a necessary part of the Great Society in which the individual no longer has specific claims on the members of the particular small group into which he was born.” Hayek, *The Political Order of a Free People*, p. 55. Emphasis added.

¹² But not entirely; a type of impartiality was invoked as a standard of acceptability of a moral theory. Gaus, *The Order of Public Reason*, pp. 294–303.

(I have argued that such would be the fate of any algorithmic rule). Note that each Member of the Public judges for herself the features of z : there is no assumption that it is a “state of nature.” Part of each person’s moral perspective is her judgment about the nature of social life with no rule over this matter. The set of options ranked as better than z by everyone is $\{a, b, c\}$, the *eligible set*. Among the remaining options $\{a, b, c\}$, all agree that $b > c$, so the “Members of the Public” can exclude c .¹³ However, society holds that a is indifferent to b , so both are in the *optimal eligible set*.¹⁴

The advantage of this social choice-inspired method is to bring individuals’ full moral perspectives to bear in ranking alternative moral rules (or interpretations of rights); the *optimal eligible set* consists of the rules that everyone can see as superior solutions to their moral reconciliation-coordination problem. Display II.4 recalls it.

		Betty	
		R_1	R_2
Alf	R_1	2 nd	4 th
	R_2	3 rd	1 st
		4 th	2 nd

Display II.4. The coordination-reconciliation game.

This game underlies the Deliberative Model: for all options above z (i.e., those in the eligible set), individuals are assumed to value reconciliation-coordination over “going it alone.” It is, of course, a purely formal model: it tells when it is rational to converge on a rule, not what rule that will be. As I have understood it in the past, the ordinal utility reflects all moral costs and benefits—all the considerations a moral agent might consider. We have seen in Part I that a sharp division between morality and interests is unrealistic (§9.3), so we should assume that the orderings are based on one’s comprehensive perspective, assuming both manifestly moral concerns and those interests that shape what moral views we accept.¹⁵ All this is consistent with our strategic model from Part I. For the sake of modeling the development of morality, this seems like the most sensible specification of the utility function. Its great attraction is that every option within the eligible set is a Nash equilibrium: if others coordinate on some rule x , one does best, *given one’s own interest-informed moral perspective*, by adopting x . This would secure a sort of moral stability.

¹³ Rule c cannot be in the “maximal set” of $\{a, b, c\}$ since everyone agrees that b is better than c . See Sen, *Collective Choice and Social Welfare*, p. 55.

¹⁴ The Pareto or Indifference rule is characterized by quasi-transitivity, not full transitivity. So $(b > c)$ & $(c \sim a)$ does not imply $b > a$.

¹⁵ It is critical to keep in mind that I am developing the strategic model of justification from Part I; this is not what a philosopher might see as a “pure” model of impartial morality.

Three concerns can be pressed about the Deliberative Model. Most obviously, there is a natural worry that in a highly diverse society the eligible set might be empty.¹⁶ The setup of the model assumes the reconciliation-coordination game, but we know that some moral perspectives are more concerned with reconciliation than others. There are actually two questions implicit in this “null set” worry. The first is that some individuals will simply prefer no morality to a moral rule on some matter. This, though, is not an objection, it is simply a possible outcome. As we saw in Part I (§10.3), climbing the ladder of impartiality may well show that moral rules over some area of social life only have a sectarian basis, and so there may well be a “null set” concerning them. Buchanan and Powell see moral progress in many “demoralizations”: “Examples include profit-seeking, lending money at interest, masturbation, premarital sex, same-sex sexual relations, interracial marriage, and (some instances of) civil disobedience.”¹⁷ All of these are the result of a “null set.” Of course, there could be an “every set is a null set” worry—that *no* moral rules can be justified at all, and we land in a state of moral anarchy. Given everything we have seen about the critical role of a shared morality in human social life in our first inquiry, this looks rather unlikely—a social life regulated by justified moral rules is basic to human moral evolution and cooperation.¹⁸ Given our strategic model (§10.3), much more worrying from the perspective of the Open Society is that a diverse society splits into different groups, with different justified rules. Thus, the real challenge of the “null set” objection to a diverse society is that there may be many moralities of many groups, with no overarching moral framework. Let us call this the *No Group-Wide Convergence Worry*.

The No Group-Wide Convergence Worry is also at the heart of the second concern about the Deliberative Model which, interestingly, comes from something of the opposite direction of the null set worry. The basic Deliberative Model confronted an equilibrium selection problem. In many cases we would expect, as in Display II.3, the eligible set to have multiple members. If so, how can society choose? Having too many choices is very much like having none: what is needed, at least for core elements (§18) of the Open Society, is a solution to the choice problem—not many, and not zero.¹⁹

Lastly, and relatedly, the social choice setup of the original Deliberative Model rather assumes that individuals simply have what we might call “inherent evaluations” of each moral rule (which is the basis of their ordering). But for many—those who stress the importance of cooperation—the moral value of a rule may greatly depend on its degree of uptake by others, while for others their moral evaluations may be far

¹⁶ See D’Agostino, “The Orders of Public Reason.”

¹⁷ Buchanan and Powell, *The Evolution of Moral Progress*, p. 56. Impartiality can spur demoralization, as more impartial arguments are prescinded from intractable disputes. See *The Order of Public Reason*, pp. 315–9.

¹⁸ Many have raised the prospect that a person might “veto” a morality, perhaps for strategic reasons. It must be stressed that on the social choice method employed by the Deliberative Model, the only concern is sincere rankings. In any event no one “vetoes” a morality; they can reduce the eligible set in a *society that contains them*. Thus we come back to the core question: if some group’s presence overly constrains the eligible set, in weighing up the costs and benefits of cooperating with them, will others converge with them on this narrower eligible set, or split away to a form a group with more attractive eligible options? This is the fundamental question for the Open Society.

¹⁹ There was some effort in *The Order of Public Reason* (chap. 7) to address this question, but it was preliminary. The following analysis builds on, and greatly develops, that preliminary model.

more focused on the inherent moral qualities of the rule (as their moral perspectives understand it). This should be included in our analysis.

§20 A Self-Organization Model

20.1. Moral Utility Functions

We can begin to see our way to overcoming all three shortcomings of the Deliberative Model by (i) integrating a basic differential valuing of reconciliation and (ii) instead of assuming a deliberative model or contractual setting, assuming that each agent acts in society based on her own reflexive understanding of what other agents are doing. Thus rather than thinking of moral justification as a centralized impartial decision for all, we shall follow the self-organization insight at the heart of the Open Society, and see whether free and equal individuals might be able to solve their equilibrium-selection problem without any agreed-upon plan.

Developing the basic evolutionary model from section 10.3, assume that each moral perspective's (or individual's) rankings can be divided into two parts. What I shall call person *A*'s (aka Alf's) *inherent* (moral) utility of rule R_1 (denoted $\mu_A(R_1)$) represents Alf's overall evaluation of R_1 in terms of how well it satisfies his reflective judgment of R_1 's inherent justice (always with some allowance for self-interest).¹ We can think of this as Alf's ranking of the rule on the supposition that he was a Kantian-like sovereign in the realm of ends—that his judgment would determine what all accepted as the morally correct rule and acted on it.² This is the way we often think about moral acceptability: we reflect on our moral perspective and decide what rule best conforms to its requirements; in our evaluation we assume a state of affairs in which a rule secures wide or general acceptance. As I said, this sort of evaluation was implicitly assumed in the original Deliberative Model's individual ranking of rules. Let us suppose that this evaluation ranges from 0 to 10 for all agents, with 0 representing a rule that the agent judges to be unacceptable from her moral perspective (no better than no moral rule at all). Of course these utilities do not support interpersonal comparisons.

Thus far, this is simply a cardinalization of the original ordinal model.³ The important development is the recognition that this is only part of the moral evaluation for those who see themselves in a reconciliation-coordination game. To be in such a game implies that, to some extent at least, one values acting in unison with others

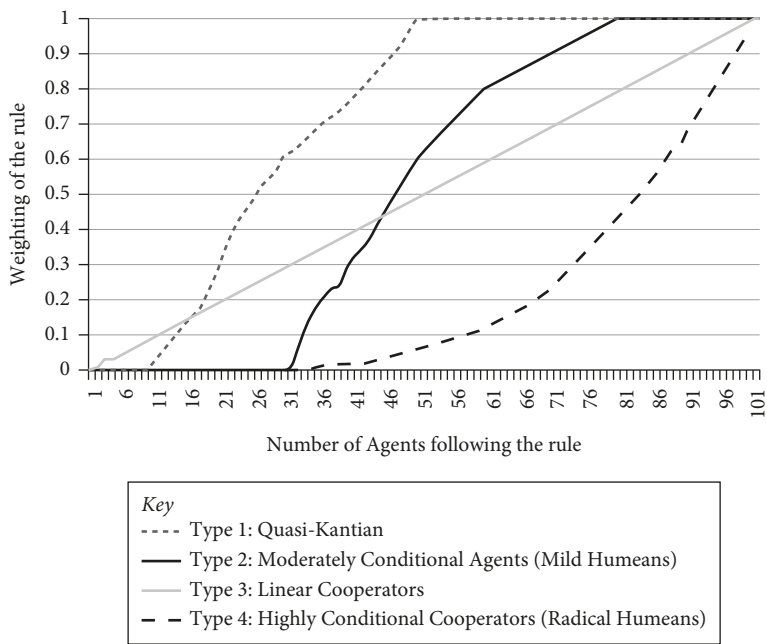
¹ It must be understood that "utility" is simply a mathematical representation of a preference ordering (a ranking in terms of choice-worthiness) that meets the standard von Neumann–Morgenstern axioms. It is not a goal or value in itself. Neither is it a stand-in for "welfare."

² "A rational being belongs to the realm of ends as a member when he gives universal laws in it while also himself subject to these laws. He belongs to it as sovereign when he, as legislator, is subject to the will of no other." Kant, *Foundations of the Metaphysics of Morals*, p. 52 [Akademie, 433–4]. Of course, like all sovereigns, ours do not entirely forget their own interests.

³ And so it requires completeness of orderings, something I did not assume in the ordinal model.

more than going it alone. The entire analysis of Part I was an extended argument for the importance of such reconciliation in human morality. Let us model this by saying that agents who recognize the importance of moral coordination also place weight on whether others share a rule. A rule R_1 where $\mu_A(R_1) = 10$, but is shared by no one, will be seen by Alf as inferior to R_2 , where $\mu_A(R_2) = 9$ but is shared by all. However, diversity again asserts itself: while to play the reconciliation-coordination game to is put *some* weight on how many others share a rule, this will differ by moral perspectives. Some place greater weight on their conclusions, as the Kantian-like sovereign in the kingdom of ends, while others put more emphasis on the Humean insight of the importance of reconciliation on common rules of morality. This is a critical difference that is at the core of our revised analyses. We thus suppose that each person has a weighting function that takes account of how many others act on a rule—how widely it is shared—and the importance to her of that degree of sharing. This weighting function for person B (aka Betty) will be denoted as $w_{B(n)}(R_1)$, which is the weight that Betty gives to Rule R_1 when n others act on it. I suppose weights vary between 0 and 1.

There are an infinite number of weighting systems. Display II.5 presents the ones on which we shall focus, here with $n = 101$.



Display II.5. Four weighting systems.

All four types acknowledge that reconciliation is part of their view of justice, and thus appreciate the social dimension of justice. “Quasi-Kantian” agents recognize some value of reconciliation: they give no weight to a rule that is not practiced by 10% of the population, but by 50% they give a rule a maximal weighting of 1. Moderately

Conditional Agents have similar shape to their weighting function but are more typical of (mild) Humean conditional cooperators: until a significant share of the population acts on a rule, they are not willing to act, and so give it a 0 weight.⁴ A rule must have 30% uptake before they give it any weight, and reaches a maximal weight at 80%. Both Quasi-Kantians and Moderately Conditional Agents, we might say, seek a moral community but not a maximally large one. Linear Agents have, unsurprisingly enough, a linear weighting function: the more who share the merrier, but as long as someone acts on their rule, they give it some positive weight. Lastly, Highly Conditional Cooperators are resolute in stressing the importance of reconciliation, and only weight rules highly when the large majority have already joined in. These Highly Conditional Cooperators are, as it were, willing to play the morality game only if most others do. Highly Conditional Cooperators give no weight to a rule unless about a third of the population follow it, and give very little weight to any rule practiced by less than 60%. They do not give really high weights until approximately 90% practice it. They are thus *highly* conditional moral agents. Highly Conditional Cooperators are in some ways the mirror image of our Quasi-Kantians, perhaps dyed-in-the-wool Humeans.

In this analysis, then, an agent is concerned with both his own evaluations of the inherent justice of a rule and reconciliation with the judgments of others, and will ultimately make a decision based on his own view of the inherent morality of the rule given his evaluative standards and the weighted number of others who are acting on the rule. Letting U_A be Alf's total comprehensive morality-based utility of acting on rule R_i , $\mu_A(R_i)$ the inherent morality-based utility of R_i given Alf's evaluative standards, w_A his social weighting, and n the number of people acting on R_i , we get:

$$\text{EQ. 1} \quad U_A(R_i) = \mu_A(R_i) \times w_{A(n)}(R_i)$$

If Alf is confronted by two rules, he will act on that which maximizes U_A . So Alf acts on R_1 rather than R_2 only if $U_A(R_1) \geq U_A(R_2)$.⁵

20.2. Model I: Fully Random

We start with a population of 101 agents, with $\mu(R_1)$ and $\mu(R_2)$ scores between 1 and 10 that were randomly assigned. Thus all agents hold each rule as an acceptable approximation of justice, if only barely (a score of 1). If an agent scores rule R_i at 0, then if that person is a member of the group, R_i is excluded from the eligible set. Perhaps if this occurs we will either go back and see whether we can climb further up the ladder of impartiality and find an eligible option or, alternatively, accept that no common morality on this matter can be secured. Cases where there is no eligible option for the entire group are largely deferred until section 21. Here I analyze a different hard case for the

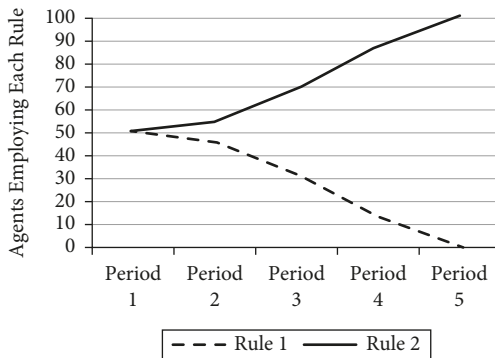
⁴ In the *Grammar of Society* (pp. 11ff.), Bicchieri models conditional cooperators as having a certain threshold, often requiring that "most" others share a rule before they will act on it. Our types do not have abrupt thresholds, but Moderately Conditional Cooperators have a significant threshold of about 30%.

⁵ And he *will* act on R_1 if $U_A(R_1) > U_A(R_2)$.

Deliberative Model—where there are multiple eligible options but none has been able to establish itself as the more popular, impartial, option in the group. We can think of our analysis as picking up where the Tooby-Cosmides-DeScoli model (§10.2) leaves off: the sides have made their cases, people have or have not changed their minds, and we are left with a certain distribution of views.

Agents were randomly assigned to one of our four weighting types.⁶ In the first period, each individual simply acts on her judgment of the inherent morality of the rule, maximizing her inherent evaluative utility (μ). In ties [i.e., $U(R_1) = U(R_2)$] an individual acts on R_1 ; perhaps R_1 is the simpler rule, and so individuals choose it in a tie. Our empirical updating rule is simple: at each period an agent calculates whether in the previous period she would have achieved more utility if she had acted on the alternative rule; if she would have, she switches in this period. She reflexively reacts to the pattern she has helped create in the earlier period.

As Display II.6 shows, Rule 2, randomly favored by 51 agents compared to R_1 's 50, went to fixation after five periods. The climb up the argumentative ladder of impartiality resulted in an essential tie, but that does not mean that society is stuck, split between two competing rules. The explanation is that valuing reconciliation generates



Display II.6. Convergence in Model I, all types.

an increasing-returns dynamic. Arthur famously showed that when a good is characterized by increasing returns—when the more others use it, the more valuable it is to any single user—the possibility of cascades arises.⁷ For all four of our weighting functions, over some range of n , the more people act on R_1 the greater weight the agent gives it—the more value she puts on it. Each of our agents is seeking to balance devotion to her own judgment of the inherent morality with some valuing of reconciliation to secure coordination. Within some range, the larger the number of others who share a rule, the greater its reconciliation benefits. While Alf may not switch from R_1 , supported by his individual mode of moral judgment, when 50% of the population are

⁶ For precise weightings, see the Appendix A. A fuller report of these results is presented in Gaus, “Self-Organizing Moral Systems.”

⁷ Arthur, *Increasing Returns and Path Dependence in the Economy*.

acting on it, as the number shrinks (that is, as others come to endorse R_2), the reconciliation benefits of acting on R_2 are apt to eventually be so great that he will abandon R_1 . Note that this is not because he had changed his relative evaluation of the importance of the two modes of reasoning, but because the value of reconciliation becomes so large as more and more others endorse R_2 .

A cascade occurred in Model I. In conformity to our strategic reasoning model (§10.3), the first agents to switch from R_1 to R_2 were indifferent in their inherent morality evaluations, such that their decisive consideration was how many others were acting on each rule. Only four agents were in this position, but this was enough to get a cascade going; so the core “quasi-indifferent” group need not always be large. After they switched to R_2 , others—who perhaps thought the inherent morality advantage of R_1 over R_2 was slightly greater than our first group (and so did not switch in period 2)—came to the conclusion that, given the slightly greater number of people following R_2 in period 2, the reconciliation benefits of R_2 now outweighed R_1 ’s inherent advantage, and so they changed in period 3 (14 agents did this). And we can see that in period 4, many of those who thought R_1 was considerably superior to R_2 now came to the conclusion that an insufficient number were acting on R_1 , and so it provided insufficient reconciliation (23 agents). By period 5, all R_1 supporters decided to endorse R_2 , including a quasi-Kantian agent (one who puts modest stress on coordination) who judged $\mu(R_1) = 9$ and $\mu(R_2) = 1$.

It might be wondered whether any specific weighting type was crucial in producing convergence, but under this same population, any three of the weighting systems (again, randomly assigned) resulted in fixation on R_2 , giving some reason to believe that the convergence dynamic is not driven by specific types. Moreover, it was typically the case that a more diverse assortment of weighting types (all four) produced convergence quicker than populations with less diversity. Combinations of types certainly can have an effect; omitting the Highly Conditional Cooperators, for example, slowed down convergence. This is interesting. In many ways, Highly Conditional Cooperators seem an impediment to moral convergence. They can be understood as viewing moral action as a Stag Hunt or Assurance Game, in which most others must play “Act Morally” before they do.⁸ One might expect them to play the “risk dominant” equilibrium. But as part of a social process, they can perform a critical role, spurring the completion of convergence, preventing it from “sputtering out.”

Finally, it might be wondered what occurs if some of the agents have no interest in reconciliation: as they understand “pure” or “maximal” integrity, it requires always acting on their Kantian sovereign judgments. Of course, given a large enough contingent of such agents, if they disagree there will be no convergence; but what if, say, 10% of the agents are of this sort? To see if such agents easily block convergence, 10% of the R_1 favoring agents in our population were replaced by such “maximal integrity agents.” We might hypothesize that these agents would either tilt the convergence dynamic to R_1 , or at least slow down the process converging on R_2 . In fact, they had no effect on the behavior of others: convergence of the rest of the population (96 agents) on

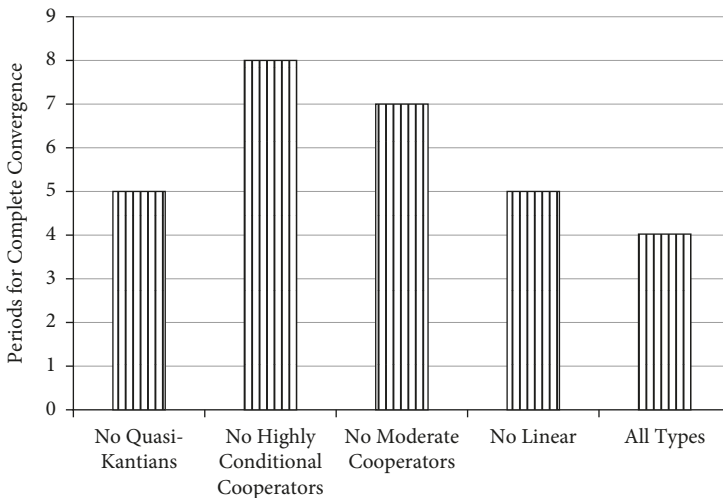
⁸ Perhaps they should be called “Rawlsians”: “Each person’s willingness to contribute is based on the contribution of others.” Rawls, *A Theory of Justice*, pp. 238, 305–6. See also Weithman, *Why Political Liberalism?*, pp. 48–9.

R_2 occurred in the same number of periods (five), with the only difference being that these resolutely non-reconciliation agents formed their own five-member R_1 network.

20.3. Model II: Moderate Polarity

We commenced with a fully random model to explore the core dynamics under conditions where the population was very closely divided, and to better see some of the effects of the different weighting systems. It certainly is clear that the dynamic does not depend on one rule having an initial overwhelming advantage (51/50 was enough). Different weightings have different thresholds and values, which help induce convergence dynamics. However, fully random distributions of inherent utility and weighting functions are hospitable to cascades, since they tend to ensure that there will be continuity of degrees of $U(R_1) - U(R_2)$ differences, such that whenever some agents switch, this will decisively affect the choice of the next agents “in line,” who then switch in the next period, and so on. The question is the extent to which a convergence dynamic applies in non-random populations with discontinuities. An especially difficult case is a polarized population, divided into two mutually exclusive groups, one subgroup thinking highly of one option and scoring the other low, with the other subgroup doing the opposite.

To explore this possibility, our group of 101 agents was divided into two “Hi-Lo” groups, one of which scored R_1 between 10 and 6, and R_2 between 4 and 1 [thus $\mu(R_1)$ are all “high,” while $\mu(R_2)$ are all “low”]; the other group was assigned the opposite Hi-Lo inherent utilities. Each polarized group had approximately an equal division of all four types of agents. The suspicion that polarization makes convergence more difficult was confirmed; very closely split [52 agents $\mu(R_1)$ high; 49 agents $\mu(R_2)$ high] polarized populations did not converge. Somewhat surprisingly, perhaps, convergence on R_1 did occur within four periods at the close but not finely balanced division [56 agents $\mu(R_1)$ High; 45 agents $\mu(R_2)$ High]. As Display II.7 shows, it was found that at this



Display II.7. Periods for complete convergence in Model II.

56/45 division, complete convergence occurred with any three types; again, the omission of Highly Conditional slowed down the process (taking 8 periods), while omitting the Moderately Conditional Cooperators slowed convergence to 7 periods. Conditional Cooperators, of course, generally put greater weight on reconciliation, and so assist in overcoming polarity. Indeed, a Hi-Lo polarized population of all Highly Conditional Cooperators with an initial 53/48 advantage for R_1 —a pretty evenly split polarized population—secured complete convergence on R_1 in four periods. Quasi-Kantians, on the other hand, tend to reinforce the split; they have their maximal weightings at around 50% of the population, and so tend to reinforce polarity when the groups are about the same size. Quasi-Kantians who find themselves in subgroups who agree with them do not easily switch rules. Perhaps the truly striking thing is that even they can be induced to leave their Hi-Lo groups and converge on a common rule, and do so when any two of the other weighting types are well represented. If Hi-Lo polarity is not too finely balanced, then, it can be overcome in a diverse population: the diversity of weighting types typically speeds up the process, inducing sufficient continuity in the populations' $U(R_1) - U(R_2)$ differences even with the population is characterized by polarized (thus discontinuous) $\mu(R_1) - \mu(R_2)$ differences.

While it is not surprising that a population composed entirely of those who put great value on reconciliation (the Highly Conditional Cooperators, for example) can overcome a polarization in a population's "sovereign moral judgments," it is, I think, worthy of emphasis that as we add diversity of weighting types, polarization can be more easily overcome than in many more homogenous populations: in Display II.7, convergence was quickest when all four types were present rather than any three, and no other population converged more rapidly than the four-type population. This could point the way to good news, for not only do we disagree about morality, but for the last 150 years Western societies have been significantly polarized between "right" and "left" justice, with the last 40 adding a number of other groups (e.g., feminists, environmentalists) who also tend to Hi-Lo judgments. Rather than reasonable pluralism, we should, perhaps, be thinking of moderate reasonable polarity. The Moderate Polarity Model gives us some reason to suppose that these sharp differences in inherent moral judgments can be significantly moderated by a diversity of weighting types. It is not necessary that we all highly value reconciliation to overcome polarization. Here, we see the possibility that one type of diversity can counteract the centrifugal tendencies of another. So far from heterogeneity always being an impediment to convergence on a shared rule of justice, some configurations of diversity can help secure agreement. This at least suggests that the issue is not "do we agree enough to live together?" but "does the overall pattern of homogeneity and heterogeneity induce convergence on common ways of living together?"

20.4. Model III: Differential Reference Groups

A simplification in the models thus far discussed is that each person takes the entire population as her reference group: in each period, her decision about reconciliation is based on what the entire population has done. But often people are concerned with

narrower reference groups.⁹ Alf might seek to reconcile his view of morality with his traditional cultural group, while Betty reconciles with those in her urban and work environments. The original Deliberative Model supposed that we seek a practice of moral accountability based on shared moral rules with the widest feasible set of other moral agents.¹⁰ But in many contexts, people might be committed to a practice of accountability—and so the shared understanding of the rules of justice on which it depends—only with those with whom they regularly interact, while others may be interested in a practice of accountability based on shared rules with some other group. In this case, the different elements of the population would have different reference groups—different groups of people with whom they seek to reconcile. Can there be shared rules by moderately polarized groups under such circumstances?

To take some initial steps in understanding the effects of different reference groups on convergence, let us analyze a somewhat challenging case: the population (increased to 150) is not only split into different reference groups, but some of the reference groups display opposite Hi-Lo polarity. In group B, three-fifths of the population have Hi-Lo bias in their inherent judgments in favor of R_1 (the other two-fifths of the group have Hi-Lo inherent justice utilities in favor of R_2), while the C group has just the opposite Hi-Lo divisions.

There seem a number of plausible hypotheses about what might happen in this more complex network. Perhaps the most obvious possibility is that convergence will occur within each group, but not between them: our polarized population models just discussed indicate that with such splits we should expect convergence on the most popular rule in each group. And of course that is what would normally happen if these are entirely unrelated reference groups, for then we simply have two independent populations. The interesting case concerns populations with overlapping reference groups. In the Differential Reference Group Model, a population of 150 agents is divided into three main groups, with two of the groups having subgroups. They are:

- Group A (50 agents): randomly divided population, not Hi-Lo (an agent may have any combination of $\mu(R_1)$ and $\mu(R_2)$ between 1 and 10; inherent justice utilities are randomly assigned as in Model I).
- Group B2 (25 agents): approximately three-fifths Hi-Lo favoring R_1 ; two-fifths Hi-Lo favoring R_2 .
- Group B1 (25 agents): approximately three-fifths Hi-Lo favoring R_1 ; two-fifths Hi-Lo favoring R_2 .
- Group C1 (25 agents): approximately three-fifths Hi-Lo favoring R_2 ; two-fifths Hi-Lo favoring R_1 .
- Group C2 (25 agents): approximately three-fifths Hi-Lo favoring R_2 ; two-fifths Hi-Lo favoring R_1 .

Note that both subgroups in B have identical evaluative utility distributions, as do both subgroups in C. The difference is their reference groups, as indicated by Display II.8.

⁹ In *Norms in the Wild*, Bicchieri extensively examines the role of reference groups.

¹⁰ Hence the worry about someone vetoing the result: it would then not be maximally wide.

<i>Group</i>	<i>Reference Groups</i>
B2: R1 Hi-Lo Biased, Parochial (25 agents)	B1, B2
B1: R1 Hi-Lo Biased, Involved (25 agents)	B1, B2, A
A: Random Group (50 agents)	B1, A, C1
C1: R2 Hi-Lo Biased, Involved (25 agents)	C1, C2, A
C2: R2 Hi-Lo Biased, Parochial	C1, C2

Display II.8. Differential reference groups.

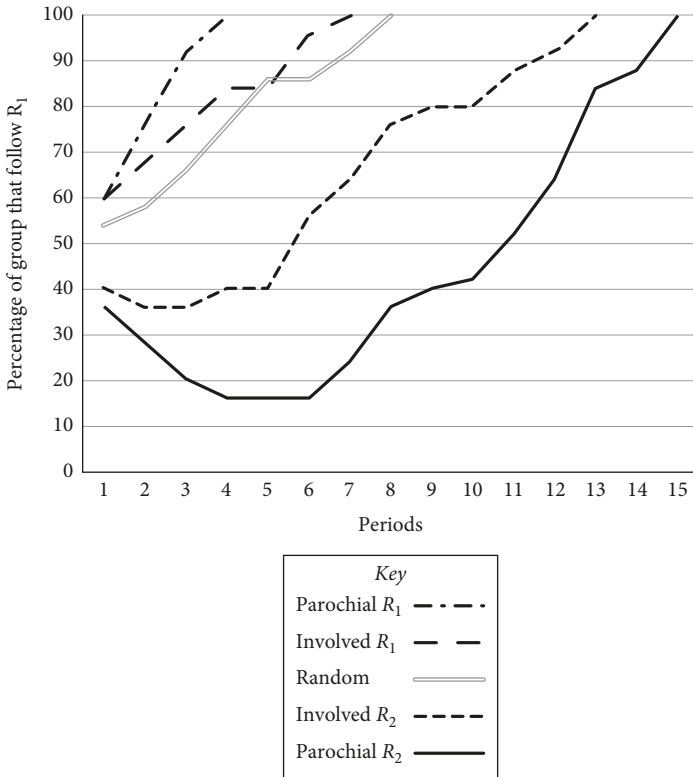
Group B2, then, is “Parochial R1 Hi-Lo Biased” insofar as they update only in relation to the choices of Group B, the Hi-Lo R1-biased group, in the previous period. This means that Group B2 (i) has a reference group of 50 agents and (ii) their entire reference group has a three-fifths Hi-Lo bias toward R_1 . B1’s reference group is 100 agents, encompassing all of Groups A and B. I call this an “Involved Hi-Lo Biased” group as it is concerned with reconciliation with its entire R_1 Hi-Lo biased group as well as the “wider” world of A. Group A, the Random group, has a reference group of 100 agents, including all of Group A itself, as well as half of both Hi-Lo biased groups (B1, which is biased toward R_1 , and C1, which is biased toward R_2). Group C is the mirror image of Group B (favoring R_2). Note that we have two parochial subgroups, B2 and C2, whose reference networks are restricted to those who share their inherent justice Hi-Lo distributions.¹¹

Display II.9 presents an enlightening simulation.¹² Agent types were approximately equally represented in all four groups. The Random Group ended up with 54% R_1 followers. So we have a picture of a society with moderately polarized and partially isolated opposing subgroups (50 agents each), and a less committed group, close to equally divided and with an entire array of judgments concerning the merits of the two rules. All in all, we might say, the ladder of impartiality has not been climbed very far in relation to the competition of our two rules.

The Parochial R_2 -biased group (64% Hi-Lo bias toward R_2 and whose entire reference group also had a 60% bias toward R_2) began, as we would expect from our analysis in the last section, by moving toward R_2 : at one point there were 84% R_2 followers. The Involved R_2 biased group (C1) was, at it were, initially pulled in two directions: some of their reference group (A) was moving toward convergence on R_1 , while C2 was moving toward R_2 , until the movement in Group A—which was part of their reference group—was strong enough to pull them toward R_1 . And, in turn,

¹¹ In these two subgroups with reference groups of 50 (B2, C2), all weighting systems were normalized so that maximum $n = 50$.

¹² This simulation was recalculated to take account of a reasonable concern advanced by Jeppe von Platz about an earlier simulation. In this case, the change in updating method (eliminating “mini-periods” in which Group A updated first) slowed down the process but did not change the overall outcome. See von Platz, <http://peasoup.us/2019/11/ppe-discussion-at-pea-soup-gerald-gauss-self-organizing-moral-systems-with-a-critical-precis-by-jeppe-von-platz/>



Display II.9. Convergence with differential reference groups.

that eventually pulled C2, the Parochial R_2 -biased group, in their wake, ending up with 100% R_1 convergence. The last to switch to R_1 were, as would be expected, Quasi-Kantians favoring R_2 in the Parochial R_2 -biased group.

Diversity of types was important, though not always necessary: convergence with only one type can occur—if we have the right sort of type. In a simulation with the same distribution of inherent utilities as in the preceding model, a homogenous population of Linear Agents, for example, failed to converge on a rule in *any* of our five groups; the same occurred in a pure population of Quasi-Kantians. In a homogenous population of Highly Conditional Cooperators, however, convergence was quickly achieved. Perhaps this should not be surprising, since Highly Conditional Cooperators give great weight to high convergence. As I said, they are dyed-in-the wool Humeans. Nevertheless, the important and surprising lesson from our simpler models is confirmed: sometimes adding *more diversity* makes agreement more likely.

Finally, *agent types* were randomly distributed in the population (up to now we have been ensuring a roughly equal representation of all agent types in each group). Here full convergence on R_1 was not achieved: 24 (of the 50) members of Group C (including all the Hi-Lo biased members of the Parochial R_2 -biased group) maintained a small R_2 network, while the rest of the population (126 agents) moved to R_1 . In Group C1

(the Involved R_2 -biased group), those with Hi-Lo evaluative utilities biased in favor of R_2 were almost all Linear Agents (with three Quasi-Kantians). As a result, those Hi-Lo biased in favor of R_2 in C1 were not sufficiently sensitive to movement in Group A to R_1 , which in turn insulated C2 (the Parochial R_2 group) from the movement in the Random group, A. Linear agents engage in such gradual updating that few could overcome their Hi-Lo bias (in inherent justice judgments) in favor of R_2 , even though there was some movement within the Involved R_2 -biased group (C1) to R_1 . Recall that in a homogenous population of Linear Agents none of the five groups achieved convergence. Again, we see how a diversity of types can generate agreement: in the R_2 -biased group there was not enough diversity of weighting types to overcome its Hi-Lo bias.

20.5. Modeling Moral Choice in a Diverse Society

All three of these models are analyses of *perfectly free and equal, moral and rational* agents, providing an initial exploration of some dynamics of rational moral choice that lead a population with somewhat stark moral disagreements to reconcile on a shared moral rule. Like social contract theories such as Rawls's, the point of these models is to understand rational moral persons and their choices, not to make predictions about actual systems of interactions.¹³ The point is not to be more "realistic" than a social contract theory: like all ethical models, these involve strong idealizations.¹⁴ The point, rather, is to capture two fundamental moral dynamics that are typically obscured in social contract theories.¹⁵

- (i) In a diverse and complex society, people not only disagree about the inherent moral merits of alternatives, but about the degree to which they seek reconciliation in a common morality. Indeed, in the third model not everyone wishes to reconcile with everyone else; *pace* the Deliberative Model, all do not seek the widest possible system of moral cooperation. In Display II.9, convergence was achieved despite the polarization and the fact that no group had an interest in reconciling with everyone (and some not even with a majority). B1 had no interest in reconciling with A, C1, or C3; B2 had no interest with reconciling with C1 or C2; A had no interest in reconciling with B1 or C1; C2 had no interest in reconciling with B1 or B2; and C2 had no interest in reconciling with A, B1, or B2. Yet, convergence was achieved.
- (ii) The models stress that, as in all major aspects of the Open Society, reflexivity drives decisions: individuals like Alf are constantly reacting to the decisions of others, which then feeds into those other's decisions, which in turn become a new consideration in Alf's decision-making. The model seeks to analyze the outcomes of decisions of free and equal individuals, with different evaluations of moral rules and different interests in reconciliation (measured in Model III

¹³ No more than the original position is a prediction about what actual people will contract. Like Rawls, I abstract from bullying, strategic behavior, inattention, and a host of other factors to better understand a core dynamic of rational agreement among free and equal moral persons.

¹⁴ See Appiah, *As If*, chap. 3.

¹⁵ Muldoon's *Social Contract Theory for a Diverse World* is an important exception.

on two dimensions: their weighting function and reference groups), reflexively interacting to produce a self-organized, and perhaps common, morality.

Highly diverse agents can converge on a common morality without any collective agreement, or informational restrictions on what they know about their moral views. A contract, bargain, or central authority is not in principle needed to adjudicate the dispute. *Self-organization thus can be recursive: it can create the rules that regulate further self-organization.* This is a fundamental result. *Not only is a social morality required to structure the process of self-organization (§18), but that very morality itself can be the result of self-organization.* We have, then, proof of concept that we need not appeal to an exogenous process such as the state or macro-selection to structure self-organization. Moreover, we have a proof of concept for what seems a striking conjecture: in some cases, diversity enhances the prospects for agreement. In political philosophy we are used to assuming that moral diversity is a problem to be overcome. Some find almost any moral diversity daunting, and so they end up modeling a social contract in which we all agree about justice. Braver souls inquire how a shared morality is consistent with some moral diversity. But we have proof of concept here for a radical idea: *increasing diversity can enhance the prospects of agreement.* In our models a diversity of weighting types often produced faster convergence than one or two types: in some cases, reducing the diversity of weighting functions halted convergence. This is not, of course, to say that diversity always makes convergence easier. At high levels of homogeneity, agreement is, of course, apt to be reasonably easy; the critical insight is that relation between diversity and tendency to converge is not monotonic. Overall our models assuage the No Group-Wide Convergence Worry (§19.3): even under what looks like fairly deep diversity (there were roughly 275 possible moral positions in our 150-agent society), the ingredients for convergence can exist.

20.6. The “Optimizing Stance”

The Self-Organization Model greatly relaxes another assumption¹⁶ of the Deliberative Model—that individuals must abandon what I have called the “optimizing stance.”¹⁷ Consider again Display II.3, which depicted the social choice problem of the Deliberative Model. An obvious worry would be if each insists on her first option (*a* for Alf; *b* for Betty; *c* for Charlie; and *d*—an option not eligible in sets for the others—for Doris). If each optimizes in the sense of insisting on their “first” option,¹⁸ then no common morality is possible. A common morality requires some readiness to coordinate on any member of the eligible set. Now it may well seem that, unless one happens to get one’s first option, one would not be acting on justice as one truly understands it. In David Estlund’s terms, for us to endorse a common rule in such circumstances, it seems that some must accept “ersatz justice”—an approximate, not true, justice—as a way to cooperate.¹⁹ In any event, it seems as if reconciliation requires that one must abandon the “optimizing stance.”

¹⁶ Besides the aim of a maximally wide reconciliation.

¹⁷ See *The Tyranny of the Ideal*, pp. 215–20.

¹⁸ They all rank “no moral rule at all” as their second choice.

¹⁹ Estlund, “The Truth in Political Liberalism.”

Note, though, that the Self-Organization Model does not require any restraint on the optimizing stance: it is a model of *optimization*—for moral perspectives that understand justice or morality to have a significant, necessary social role. On these views, the point of a commitment to social morality or justice is not simply to have a true, correct, or most reasonable belief about inherent morality (morality from the point of view of the Kantian-like sovereign), but to realize moral or just social relations in one's life. On the Self-Organization Model's extended utility function, that is part and parcel of morality and justice. As Kant himself stressed, in a state of nature without common rules of justice, even if we imagine individuals "to be ever so good natured and righteous," when each does what "seems just and good to him, entirely independently of the opinion of others" they fail to secure justice.²⁰ The Self-Organization Model builds on this insight, allowing for a range of different commitments to this understanding of justice and morality, from our "Quasi-Kantians" to our radical Humeans. None of them abandons the "optimizing stance": they optimize reflexively, given the optimization of others.

However, to participate in a wider cooperative morality, those who do *not* uphold such social-realization-sensitive moral perspectives typically must indeed abandon the optimizing stance. On moral perspectives where "optimization" simply means acting on one's best understanding of *inherent* morality or justice (optimizing, assuming one is the Kantian sovereign), when others do not share one's view of inherent justice one must either forgo moralized social relations (simply doing what "seems just and good to him, entirely independently of the opinion of others") or find reasons to follow an approximate or "ersatz" view of injustice/morality.²¹ Commitments to the value of social cooperation, moral accountability, social trust, civic friendship, community, or a principle of "meeting others halfway" may all induce individuals to endorse and act upon moral rules that do not conform to the best option in their eligible sets (based on inherent justice rankings). Moreover, if such allied values are present, societies no doubt will more easily overcome polarities of judgment. As we saw in some of our simulations, polarized societies can gravitate to two stable networks. Invoking these other values may induce convergence in these more divisive circumstances.

§21 Moral Diversity in the Open Society

21.1. The Final Element of the Complex Society Moral Package

The models explored in the previous section partially assuage the No Group-Wide Convergence Worry. Great diversity does not preclude convergence. As Appiah notes,

²⁰ Kant, *The Metaphysical Elements of Justice*, p. 116 [§43]. Emphasis added. Perhaps our "Quasi-Kantians" are really simply Kantians after all.

²¹ To overcome what I have elsewhere call "the Gap" between what one thinks is best and what others will accept. See "The Commonwealth of Bees."

“coexistence ... depends on being able to agree about practices while disagreeing about their justification.”¹ The existence of an Open Society requires that there be great convergence on its basic moral elements (the Complex Society Moral Package, §18.5). This is important—empirical studies indicate that a diverse Open Society requires a shared “generalized morality in which abstract principles or rules of conduct are considered equally applicable to a vast range of social relations beyond the narrow circle of personal acquaintances.”² In contrast, “limited-group morality” is apt to reinforce parochial outlooks.³ Nevertheless, there will not be society-wide convergence on many moral rules. Even in our models we saw that in some cases agreement does not occur: groups gravitate to different rule networks on some issues. This will obviously be the case when they do not share an eligible set, but polarity of evaluations within the eligible set can also produce such separation. This is not surprising: as we saw in Part I, when no candidate rule is able to sufficiently climb the ladder of impartiality, such separation is a possible outcome (§10.3).

Now when it is recognized that both rules are in a shared eligible set, living with such difference should, in practice, not be unsettling. Each of the rules functions as a subgroup’s coordination convention. There is always some arbitrariness in the rules we live by, and mere difference here need not be divisive.⁴ Matters become considerably more perplexing when the groups do not share a common eligible set: group A coordinates on a rule that is not seen as eligible by group B. In non-moral cases—say, mere taboos or customs—this too need not be divisive. “That’s their odd way, this is our more sensible way—live and let live.” We can thus appreciate the attractions of moral relativism in accommodating diversity: moral disagreement becomes mere difference, and there is no reason why animosity should arise. We need not reject moral relativism here—certainly some perspectives will assert it—but neither should we rely on it.⁵ Most moral perspectives are not relativistic; moreover, we have seen that commitments to objectivity have an important function in human moral life (§10.2).

Hence the difficulty: how to reconcile a commitment to the objectivity of one’s moral beliefs with the fact that in a diverse society other groups of individuals will converge on some beliefs that one does not deem eligible. Now if what they are doing is harmless, once again no great problem presents itself: “Muslims believe that they ought to make the hajj and Catholics that they ought to go to Mass. But if you don’t have the beliefs that give those acts their meanings, you presumably think that the people who do think so are mistaken. ... Of course, you probably don’t think there’s much harm done if people do go to Mecca. They think it’s right. We don’t. We don’t think it’s wrong, either, though.”⁶

What, though, if you do think them wrong? The United States Conference of Catholic Bishops believes gay marriage is wrong; you believe it is a right. Alf eats meat and Betty believes this to be morally wrong. In such cases we have not simply failed

¹ Appiah, *Cosmopolitanism*, chap. 5.

² Platteau, *Social Norms and Economic Development*, p. 304.

³ *Ibid.*, pp. 311ff.

⁴ See Vanderschraaf, *Strategic Justice*, pp. 67ff. A related possibility is that the different rules are simply different applications of the same moral principle, adjusted for local conditions.

⁵ For a sustained criticism of this solution, see Appiah, *Cosmopolitanism*.

⁶ *Ibid.*, chap. 1.

to coordinate on a common morality: we are convinced that the others have coordinated on the immoral. It would seem that a commitment to moral objectivity must entail that we censure them. Those who have done wrong are to be held accountable and disapproved of. We can be accepting of difference, but the range of tolerance for immorality is severely restricted.

The rise of religious toleration in Western Europe and intergroup toleration in multiethnic cultures such as India suggest a rather different conclusion. Western Europe in the early modern era saw the rise of multi-confessional states in which individuals, devoted to the absolute truth of their doctrine and its moral injunctions, engaged in non-censorious social relations with those of different faiths. Toleration was not the victory of secular reason over sectarian belief: it usually arose as a practice of coexistence among those with opposing sectarian beliefs.⁷ Neither was toleration secured by reconceptualizing religio-moral beliefs as simply group taboos or injunctions limited to the faithful: the doctrines typically contained rich moral-universalistic codes regulating speech, family life, sexual conduct, charitable obligations, and so on. Yet many saw their neighbors of other faiths as good-willed and trustworthy partners and did not censure them for failing to live up to their religious morality. For example, religiously divided Dutch towns and neighborhoods explicitly sought to provide the conditions for fruitful coexistence; a typical statute forbade neighbors “In any way to offend, injure, or speak any contemptuous words . . . to one another, or . . . do any violence or make any threats.”⁸ Today, too, many people’s understanding of universalistic morality goes far beyond the WEIRD morality of freedom, fairness, and harm, and many of these non-WEIRD rules are seen as moral, not conventional or parochial: living with each other requires similar feats of tolerant coexistence.⁹

True moral diversity is compatible with low levels of conflict when a subgroup’s judgments of moral truth do not ground a practice of holding outsiders accountable for violations.¹⁰ We have seen (§§9.4, 10) the importance of a practice of accountability for upholding moral coordination within a group; calling people out, excuses, and censuring are all part and parcel of a system of moral cooperation. But we have also seen that when a person does not believe that she has done wrong, seeking to hold her accountable is apt to invoke resistance and counter-punishment. After all, as she sees it, she has done nothing wrong, so you are the aggressor: members of a group who do not see a rule as grounding legitimate (i.e., justified) moral demands are apt to resist, and to feel rancor when others censure them by appeal to it. Consequently, if a practice of accountability is to function as a device of cooperation rather than an

⁷ For a comprehensive study, see Kaplan, *Divided by Faith*.

⁸ Quoted in *ibid.*, p. 251.

⁹ Survey research indicates clear moral differences throughout the world. See, e.g., American-Western European Values Gap, PEW RESEARCH CENTER (Feb. 29, 2012) <http://www.pewglobal.org/2011/11/17/theamerican-western-european-values-gap/> [<https://perma.cc/U3HX-82JG>]; Shweder, Mahapatra, and Miller, “Culture and Moral Development”; Haidt, *The Righteous Mind*, pp. 95ff.

¹⁰ According to Akeel Bilgrami, the source of tolerance in Gandhi’s philosophy is “to be found in his conception of the very nature of moral response and moral judgement. The ‘satyagrahi’ or non-violent activist has to show a certain kind of self-restraint, in which it was not enough simply not to commit violence. It is equally important not to bear hostility to others or even to criticize them; it is only required that one not follow these others, if conscience doesn’t permit it.” “Gandhi, the Philosopher,” p. 10. My thanks to Ritwik Agrawal for enlightening discussions on this matter.

impetus to conflict, the moral rules being upheld must be publicly justified (§9.3). In cases where a group has truly controversial moral beliefs, it follows that these cannot be robustly justified to outsiders. It is precisely in these cases—*moral belief without public justification*—that accountability has dysfunctional features. And, so, it is in these cases that breaking the link between a person doing wrong (as you see it) and you holding that person accountable is necessary for the moral life of the Open Society.

To be sure, there will always be limits. There are some actions that one perceives as so morally hideous that one simply must censure those who engage in them, whether or not the morally correct rules can be justified to them. It is these cases that are most apt to immediately spring to mind, and of these cases abortion is the contemporary exemplar. Such cleavages can create hostility and even hatred between the opposed groups (§10.4). If they predominate, moral diversity will render the Open Society unstable, perhaps impossible, and certainly acrimonious. Highly diverse societies thrive when, only with great reluctance, people censure others for failing to act on moral requirements when those requirements are not publicly justified. Thus we come to the final (and a crucial) element of the Complex Society Moral Package: *the necessity of a practice of accountability based on publicly justified rules*.

This element has been overlooked. As I pointed out at the outset of this Part, too often the idea of the Open Society has been associated with a Millian secular universalism. On this view, the Open Society is to be based on a progressive consensus on the truth of WEIRD morality and the relegation of other normative belief systems to local taboos, mores, social norms, conventional rules, or other cultural artifacts. If the advocate of the Open Society rejects the necessity of public justification for a system of accountability, it is hard to avoid this Millian vision. The obvious reasoning goes thus: “if people hold their moral beliefs to be objectively correct, those who disagree are objectively wrong, and so when this leads to immorality they should be condemned.” If there is such a direct and necessary link between moral judgment and accountability, moral diversity must result in constant conflict and dissension. For the Open Society to free itself from commitment to this progressivist utopia (or, I suspect, dystopia), while also avoiding retreat into public moral relativism, it must embrace public justification—not for one to believe that something is objectively correct, but to employ that belief in a practice of accountability.¹¹

21.2. Aggressive Individual Moralism

I have been critical of claims about our “tribalism”—they catch some, but miss a great deal, of our complex social nature. Joshua Greene writes not just of “tribalism” but of “moral tribes.” He provides a fable of two tribes, “engaged in bitter, often bloody conflict.... They fight not because they are fundamentally selfish but because they

¹¹ The view articulated here, I believe, is a version of what Appiah calls “cosmopolitanism,” though one that I believe accepts rather deeper moral diversity than on his analysis. See his *Cosmopolitanism*. For a defense of the intimate relation between public justification and a practice of accountability, see Gaus, *The Order of Public Reason*, chap. IV.

have incompatible visions of what a moral society should be.”¹² Greene’s “moral tribes” amalgamate three distinct phenomena: that humans tend to form cooperative groups bounded by kin, ethnic, national, or other markers; that internally a cooperative group is structured by a shared system of morality; and that their moralities lead them into bitter conflict. The amalgamation, useful for Greene’s tale, obscures a critical point: aggressive morality need not be about self-defined groups at all. Indeed, one of the features of modern moral thinking over the last 200 years has been a highly individualized form of moral thinking that readily leads to bitter conflict.

The broad set of moral theories I have in mind are, roughly, characterized by a picture of moral judgment in which a competent moral agent deliberates on “impartial evidence” and soundly concludes that it is a moral fact that “ ϕ is wrong,” where the beliefs of others about whether ϕ is wrong cannot be truth-makers for “ ϕ is wrong.” Although an epistemically modest person will take account of other people’s deliberations as evidence (as would a scientist), such social facts cannot make a moral claim true or false, any more than facts about what people believe is the distance to the moon are truth-makers for the “the distance from the earth to the moon is 238,900 miles.”¹³ In addition, on this individualist view a person who soundly concludes that “ ϕ is wrong” has strong reasons to demand that others refrain from ϕ -ing and to censure, criticize, and perhaps punish them if they do not.

In the past fifty years, some have resisted this picture—Rawls, Strawson, Baier, Gauthier, and Hayek chief among them. In one way or another, those in this latter tradition argue that moral truth itself is in some way a function of a type of moral agreement, consensus, or public justification. Theirs are social theories of morality: while morality by no means reduces to social facts, certain social facts (or counterfactual social claims) are a necessary part of the grounding for a claim that an action is right or wrong. In this, such theories display continuity with the Modern Egalitarian Moral Package, which expressed an irreducibly social understanding of moral life. In contrast, the individualist moral view can warrant a competent deliberator judging the acts and practices of others as wrong “entirely independent” of anyone else’s judgment, simply on the basis of her own conscientious deliberations.¹⁴

Having disaggregated Greene’s “moral tribes,”¹⁵ we can now see that a major source of moral conflict is hyper-individualized moral judgment. On the basis of their own deliberations, hyper-individualists see themselves as warranted in condemning their fellows. As society becomes increasingly complex and new actions and ways of living arise at ever-increasing rates, wrongness too abounds. Those who are ahead, behind, or out of sync with the individual’s conscience are deemed immoral and worthy of social censure. Of course the hyper-individualist insists it is not *her* deliberation that she demands be respected: it the moral facts that speak through her.¹⁶ But she need not listen to those facts with others: she hears their voices alone.

¹² Greene, *Moral Tribes*, p. 5.

¹³ It is important to stress that all sorts of social facts can, of course, enter into the content of these moral judgments (“polygamy is wrong,” “divorce is permissible”).

¹⁴ See further my “Private and Public Conscience.”

¹⁵ It must be acknowledged that Greene by no means always sticks to the “tribal” story.

¹⁶ As Mackie noted, in some sense it is often claimed that these facts do speak to us, instructing us what to do. J. L. Mackie, *Ethics: Inventing Right and Wrong*.

Some version of this general hyper-individualist approach is dominant in moral and social philosophy, in the university generally, and is part of the contemporary *Zeitgeist*.¹⁷ Criticizing others for immorality or injustice is extraordinarily easy: no public and agreed-upon violation need be required. And because condemnation and censuring are seen as direct implications of moral judgment, censuring and counter-censuring explode. All this is aggravated by social media, which drastically lowers the cost of censuring. While in some instances this can allow a moral rule to be far more effectively enforced (as the “#MeToo” movement demonstrated), it also means that censuring others for failing to live up to the demands of one’s conscience has become an extraordinarily low-cost activity. No need to confront them and risk an unpleasant encounter (think again of Cephu) and perhaps incur counter-punishment: we can costlessly punish on our smart phones. Like-minded individuals combine to mob perceived offenders with ease, indeed with enjoyment.

None of these things spells the death of a tolerant, morally diverse Open Society. But they do remind us that constant talk of “tribes”—with its implicature about the dangers of being “primitive” rather than enlightened modern individualists—blinds us to the ways in which modern hyper-individualism in moral thought is at least as much a source of bitterness and conflict. A mob of high-minded individual moralists, attacking a morally opposed view in a university protest or on social media, is not a tribe. To say so is to insult the subtle and cooperative life of tribal societies.

21.3. Looking the Elephant in the Eye

The elephant in the room, which I have thus largely ignored, is whether there is any sound basis for all these moral judgments in a complex society. Recall Hayek’s basic challenge (§2): given that our complex society has evolved, and thus we cannot know the purposes our moral rules serve, how can we possibly justify them? We now have high enough ground to look the elephant straight in the eye.

Hayek’s problems are not precisely ours. Hayek, we have seen, adopts a version of macro-selection to explain the functionality of the complex system of the Open Society (§§11.4, 17.3). Its rules have evolved in competition with those of other orders—consequently, the functions of its rules are generally causally opaque to us, just as the function of our genes is generally opaque. Given this, for the most part Hayek urges deference to the traditional, evolved, moral rules.¹⁸ They cannot be justified in terms of their purposes or functions, being the result of an evolutionary process. It is difficult for a macro-selection account to avoid this conclusion: if there is strong macro evolution, the adaptiveness of the system depends on following the selected rules. However, I have argued that in the current environment the case for the macro selection is underwhelming. The self-organization account models the evolution of complexity without a clear selection mechanism: reflexive adjustments in the

¹⁷ This can be understood as an extreme version of the “privatization of morality,” which on Weber-inspired accounts is critical to capitalism. Platteau, *Institutions, Social Norms and Economic Development*, p. 302.

¹⁸ Hayek, *The Fatal Conceit*, pp. 66–71.

autocatalytic process of niche creation are constantly opening new possibilities at an increasing rate, and individuals' understandings of the moral rules they can endorse are part of this process. But this makes the very concept of a moral rule's "function" or "purpose" obscure. To be sure, moral rules serve a very general function of providing settled expectations, and consensus on prohibitions (as well as some permissions) that in one way or another permit individuals to coordinate their plans. More generally, in Part II we have seen that the Open Society requires a Moral Package if self-organization is to lead toward extended cooperation rather than conflict and division. Nevertheless, it will almost always be impossible to identify the specific purpose of a given rule. To the extent that morality is self-organized, any given rule is a vector of innumerable purposes and choices, each one of which may or may not have had some purpose or aim in mind. Moreover, there is no assumption that individuals within a self-organized moral system concur on the self-organization analysis. The diversity of moral perspectives (§16) remains fundamental to the moral life of the Open Society. Consider, for example, just a few of the perspectives that populate the Open Society.

Consequentialists: Some will always insist that justified moral rules must seek to maximize some overall metric, such that rules are justified if and only if they are effective in producing favored social states. Even Hayek suggests that sensible people are attracted to utilitarianism because it is a basis for rational examination of existing rules.¹⁹ We might distinguish three sorts of consequentialists. *Cognitively Open Consequentialists* will be continually evaluating their favored rules to determine whether they are producing the expected consequences. As we are about to see in Part III, in the swiftly changing landscape of an autocatalytic, reflexive, and complex society, the unexpected and the unpredictable are the norm: the project of pursuing an overall, systemic goal for society is bound to disappoint. We have seen, moreover, that there are no algorithms for producing valued social states via rules (§19.2). Hence we can expect Cognitively Open Consequentialists to be continually changing their ordering of rules as new information comes in: they will be a constant source of diversity of moral judgments, though if their numbers are too high, they can be highly destabilizing. In contrast, *Dogmatic Consequentialists* will be very reluctant to change their opinions about the total social consequences of rules.²⁰ Again, as we are about to see in Part III that, given the absence of data and the difficulty of evaluating it, this will often be a plausible position ("you can't be sure rule *R* didn't work; after all, favorable condition *x* could have been brought about by it"). As consequentialists lean toward the dogmatic end of the scale, they tend to be conservative and stability-enhancing. *Myopic Consequentialists* will justify rules in terms of very specific and limited purposes ("Does this rule increase economic growth?"). These judgments can be accurate in the short-to-medium

¹⁹ Hayek, *The Mirage of Social Justice*, pp. 17–8. This by no means makes Hayek—as some have suggested—a sort of rule utilitarian. For a subtle treatment, see Mack, "Hayek on Justice and the Order of Actions."

²⁰ There is good reason to think that these are two distinct cognitive types—what Philip Tetlock calls cognitive "foxes" and "hedgehogs." See §26.2.

term (§26.2): the myopic focus means that unanticipated effects of the rule are largely ignored.²¹

Contractarians: Like Rawls and Gauthier, some will hold that diverse individuals would all agree to a set of rules of justice and institutions, some of which may provide for relatively clear articulations of required patterns or outcomes. Insofar as they do, they may divide into types similar to consequentialists (open, dogmatic, and myopic).

Transactional Fairness: We have seen that humans are generally averse to domination and bullying. Rules that express the basic equal status of agents thus will be highly attractive and relatively stable when equality and fairness are understood in terms of features like the reversibility of roles (a person can endorse a rule regardless of the position she occupies), the absence of manifestly subordinate roles, and so on. Those stressing respect for others are instances of this type. On the other hand, to the extent that such transactional equality depends on background conditions (say, the degree of wealth inequality in a society), transactional fairness judgments may be subject to more variation as these conditions change.

Deontic Views: Some will base their judgment of the inherent morality of the rule largely divorced from any projected consequences. In some places Hayek seems to adopt this view insofar as he insists that we should be “dogmatic” in defense of liberal principles and make “no concession to expediency.”²²

Traditionalists: At other times, we have seen, Hayek argued that given the great uncertainty of the Open Society, a deference toward accepted moral rules is appropriate. While I have been skeptical of his group-selection argument, many may nevertheless believe that rules that have survived for a reasonably long time have shown themselves to be robust in the face of social change, and thus merit deference.

Conformists: Conformists may act like traditionalists at times, but we would expect them to be much quicker to join cascades of change. For example, in the 1980s both traditionalists and conformists may have opposed gay marriage on similar grounds, while in the current decade conformists, but not traditionalists, may have switched to supporting gay marriage. As we saw earlier (§11), conformists are critical for culture, as they are for moral stability. Too much political philosophy finds moral stability mysterious because it assumes a world of autonomous trendsetters. Yet if we are all traditionalists, moral change is indeed mysterious. A diversity of types, of course, helps explain how we can have both stability and change,²³ but, interestingly, conformists can be critical forces for both—holding on to currently accepted rules but joining change cascades as they proceed.

²¹ Ignoring the unseen thus may be a strategy for stable justifications in a highly dynamic society. Cf. Mack, “What Is Seen and What Is Not Seen: Hayekian Extensions of Bastiat’s Insight.”

²² Hayek, *Rules and Order*, p. 61.

²³ See Gaus, *The Tyranny of the Ideal*, pp. 223ff.

To be sure, this is just a sampling of possible types, and within each type people will have sharp differences in their judgments. The point is that this diversity is ineliminable in a free society. We disagree about the reasonable basis for justification. To be sure, some bases may be more in tension with our analysis of the Open Society than others, but an Open Society cannot suppose that everyone accepts a common view of morality and justification. Moral diversity is not a condition to be overcome: we are not all on our way to becoming Millian utilitarians, Hayekian traditionalists, or transactional egalitarians. This is the critical difference between the liberal progressivist view of the Open Society and that presented here: any supposition of consensus on moral perspectives is inconsistent with the very nature of a free and open society. That is the great insight that drove Rawls's political liberalism project: diversity of opinion is the natural outcome of life under free institutions.²⁴

Yet that does not block the justification of a common morality. In dynamic contexts, the ability of a candidate rule to secure approximate robust public justification—a justification that withstands argument and critique (§§9.3–4)—is essential in recalibrating the diverse order on a revised shared framework. Because the Open Society is characterized by increasingly diverse perspectives, successful public justifications must be as accommodative to diversity as possible. A social morality that claims to be “publicly justified” within a narrow public (say, reasonable secular liberals) will not secure widespread conformity through widespread sensitivity, as it excludes a wide variety of perspectives from the justificatory public. Whatever normative aims such a theory may secure, it cannot be the normative basis of the Open Society. Only by the oppressive use of social or state power²⁵—power that cannot be justified to those over whom it is wielded—can such a morality be imposed on our increasingly diverse societies. The great insight of the public reason tradition—bolstered by our Self-Organization Model—is that this does not preclude a justification of the Open Society. As we have seen in the limited context of our model, a variety of different moral perspectives can, counterintuitively, enhance the ability of a society to secure public justification on shared moral rules. Each has their own idea of the point and value of these rules, yet each can participate in, and indeed enhance, a social process that can generate a self-organized social morality.

²⁴ Rawls, *Political Liberalism*, p. xviii.

²⁵ As Mill stressed in *On Liberty*, the social power of informal sanctions can be even more oppressive than state power.

PART III

THE COMPLEXITIES OF SELF-GOVERNANCE

Can the Open Society Be Governed?

The first two Parts have responded to two of Hayek's unsettling claims concerning morality: that, given our moral evolution, we may be unfit for the Open Society; and that the Open Society is so complex as to befuddle attempts at moral justification. Each of these turned out to indeed be pressing problems, yet in both cases the resources of humans in a complex society are richer than Hayek thought. In the course of responding to the second unsettling thesis, I have presented a somewhat novel understanding of the Open Society and justification within it. In this third inquiry I turn to Hayek's last unsettling thesis, and that which has been the focus of greatest criticism: that our complex Open Society is in many ways beyond human control and governance. It is impossible to evaluate this claim until we specify it more carefully. Before we begin to look at problems of complexity, we must first better understand self-governance and its dimensions.

§22 Self-Governance

22.1. Self-Governance and Self-Organization

In the previous Part, I presented an analysis of the autocatalytic complexity of the Open Society. Possible niches arise faster than they can be filled, and as new niches are filled, ever more possibilities arise. As Kauffman's Principle of the Indefiniteness of Functional Searches (§16.4) implies, the space of the possible is unbounded. Which of the indefinite expanding possibilities will come to fruition is unpredictable, but given the path-dependent nature of evolution, which ones eventuate affect what niches appear later. And given reflexivity, we are always responding to new patterns we have just produced—and to those responses. This has led us to reconceive the division of labor, as not simply dividing up tasks and inducing specialization, but as ways for individuals to search out possibilities that can attract others into novel mutually beneficial cooperative relations, which are then woven together in unexpected ways. This ever-expanding and changing web of diverse cooperation is exactly what we ought to expect from an ultra-social species whose social existence is built on reciprocity and cultural imitation rather than altruism or rational self-interest (§8). As I stressed in Part I, we are neither simply Machiavellian strategists nor thorough cooperators, but are constantly on the lookout for cooperative possibilities.

As I argued near the close of Part II, neither moralistic hyper-individualism nor top-down social contracts are suitable for morally structuring the Open Society's self-organizing complexity. Indeed, I have proposed that the most plausible approach to social morality in the Open Society conceives of it as a self-organizing moral framework that structures the ongoing self-organization of a highly diverse and evolving social world. All of this, I have argued, can be understood as a process of bottom-up public justification. In contemporary political philosophy, many critics see public justification as a rarefied academic doctrine, focused on obscure concepts and implausible empirical claims. And all too often it is. I have tried to show, though, that properly understood it is a real part of the practice of justification and argumentation that has evolved along with humans.¹

If, however, the Open Society depends on constant reflexive self-adjustment of individuals seeking out beneficial niches, we might wonder whether there is room for self-governance. Can we govern the self-organized Open Society? Jenann Ismael distinguishes self-organization from self-governance:

Self-organizing systems are a special subset of dynamical systems. The hallmark of self-organization is the emergence of order from the interactions among a typically large number of components without any centralized control. Central examples include termite colonies, schools of fish, unregulated crowds, and free market

¹ Kevin Vallier is noteworthy in stressing evidence for the real social role of public justification. See his *Must Politics Be War?*

economies. In cases of pure self-organization there is no real centralization of information or control, but the behavior of each affects that of the others in a manner that produces an overall appearance of deliberately coordinated activity.²

In contrast, self-governing systems:

are systems in which at least some organized activity is the result of a centralized process that involves the sharing of information and the formation of an overall plan and deliberate coordination of joint activity. . . . In a self-governing system . . . at least some of the information distributed throughout the systems is collected, synthesized, and used to fuel a decision procedure that plays a role in guiding the system's behavior.³

As Ismael characterizes them, self-organization and self-governance are not incompatible: self-government requires that there be a "controller" who collects *some* information at the system level, and then can use that information to "fuel a decision procedure that plays *a role* in guiding the system's behavior." My aim in this Part, then, is not to debunk the possibility of self-government, but to investigate what types and levels of self-government are consistent with the self-organization of the Open Society. For political philosophers to endorse overly ambitious commitments to democratic self-government—ones that are incompatible with the self-organized nature of social life in the Open Society—is not a democratic virtue. It is a recipe for ongoing disappointment and eventual distrust of self-governance (too often leading to temporary reversion to rule by the Big Man, a possibility always on our horizon). If our plans and dreams for self-governance go unfulfilled, surely someone—*hoi polloi*, the capitalists, the politicians—must be to blame. My concern in this inquiry is to better understand in what ways self-governance can be effective in a complex order, and when ambitions of control are delusionary and dangerous.

22.2. Governance Contrasted to Legislation

Political philosophers commonly embrace what I shall call the Legislative Theory of Democratic Self-Governance, consisting of two claims:⁴

- (i) *Popular Control*: In a democratic polity the people "achieve control over the state" first and foremost by exercising broad control over legislation.⁵
- (ii) *Legal Centralism*: Legislation is the primary mode by which a society regulates social relations, and it influences behavior primarily through the threat of punishment.⁶

² Ismael, *How Physics Makes Us Free*, p. 19. I thus depart from those who consider self-organization a type of self-governance. See Volker, "Governance and Complexity."

³ Ismael, *How Physics Makes Us Free*, pp. 19–20. See also her "Self-Organization and Self-Governance."

⁴ This section draws on Barrett and Gaus, "Laws, Norms, and Public Justification: The Limits of Law as an Instrument of Reform."

⁵ Pettit, *On the People's Terms*, p. 187; Waldron, *The Dignity of Legislation*, esp. chap. 1; Waldron, *Law and Disagreement*, esp. chaps. 1 and 2.

⁶ See Ellickson, *Order Without Law*; Mackie, "Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience."

Popular Control seeks to ensure that the people control legislation, and Legal Centralism holds that governance proceeds primarily through legislation; laws are effective to the extent that those who would otherwise engage in unlawful behavior are deterred by coercive threats. Together they sum up a widely endorsed understanding of democratic governance of society.

Legal Centralism implies that the scope of self-governance is wide-ranging because the scope of legislation is, and under Popular Control legislation is simply a matter of the people ultimately deciding who legislates. Legislation is determined by a choice of the legislature, so governance is largely a matter of what decisions it will make. If informal moral norms seem to be hindering what society wishes, or if society does not seem to be self-organizing on the correct norms, law can effectively intervene to set things right.⁷

Now of course (as we shall see in this Part) legal and political institutions are necessary for innumerable aspects of large-scale cooperation. No sane advocate of the importance of self-organized social life would deny that. The question is whether these formal institutions control, supplant, supplement, or interact with the basic framework of social rules and moral convictions. Increasingly, it is coming to be recognized that Legal Centralism is implausible: legal and political regulation without an underlying social normative framework is ineffective, and the law cannot dictate this framework. Gerry Mackie has pointed out that there are hundreds of critical cases around the world in which practices—among them female genital cutting, caste discrimination, child marriage—have been widely criminalized yet continue to be practiced. Laws that depart from the basic moral and social norms of a society mostly likely will be ignored, often engendering contempt for the law. As Mackie, following Iris Marion Young,⁸ concludes:

Criminalization is an appropriate response to a criminal injustice, a deviation from accepted norms, its harmful consequences intended, knowingly committed by identifiable individuals, whose wrongdoing should be punished. It is not an appropriate response to a structural injustice, in compliance with accepted norms, its harmful consequences unintended byproducts, and caused by everyone and no one. The proper remedy for a harmful social norm is organized social change, not fault, blame, punishment.⁹

In recent years, students of social change have come to something of a consensus that effective legal regulation cannot stray too far from the underlying informal social rules.¹⁰ One of the most striking “social experiments” based on this insight was that of Antanas Mockus, mayor of Bogotá in the late 1990s and the first decade of the 2000s.¹¹ Mockus’s aim was to harmonize legislation with social morality; he recognized that

⁷ See Sunstein, *How Change Happens*, chap. 3.

⁸ Young, *Responsibility for Justice*.

⁹ Mackie, “Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience.”

¹⁰ In addition to *ibid.*, see Bicchieri, *Norms in the Wild*; Bicchieri and Mercier, “Norm and Beliefs: How Change Occurs.”

¹¹ For a short description of this experiment, see Antanas Mockus, “Building ‘Citizenship Culture.’” See also the essays in *Cultural Agents Reloaded*, edited by Carlo Tognato.

unless supported by the underlying informal moral and social framework, attempts to induce change though law would not succeed. For example, Bogotá was characterized by a very high rate of traffic fatalities in the mid-1990s, with widespread disregard for traffic regulations. Mockus distributed 350,000 “Thumbs Up/Thumbs Down” cards that drivers could display in response to dangerous driving by others, to drive home the message that such behavior was not only illegal, but violated the informal normative judgments of other drivers. Along with related programs, Bogotá witnessed a 63% decrease in traffic fatalities between 1995 and 2003. Similar programs based on harmonizing the law with informal social normative expectations led to decreases in water usage and, critically, homicides.

In lieu of an informal moral framework that coheres with the law, in a wide variety of cases (including traffic laws, which look like simply pure coordination games) we cannot expect the mass of citizens to conform unless coerced by high penalties. And in the absence of a moral framework, we cannot expect those occupying positions in the formal institutions (in charge of administering those penalties) to be guided by its rules, rather than taking advantage of the myriad opportunities for enriching themselves.¹² The European Union is currently witnessing the consequences of assuming that the rule of law is only about formal institutions and rules, as politicians in Hungary increasingly ignore these rules in pursuit of nationalist and illiberal legislation. And private property institutions designed to promote cooperation can—and very often do—lead to kleptocracy.¹³ Without compatibility with the self-organized social morality, law and politics become simply additional devices by which some use power to extract from others.

Laws that run counter to the moral norms of the populace are not only apt to be ineffective but, according to William J. Stuntz, self-defeating: the very enactment of a controversial law that runs counter to moral norms often strengthens, not weakens, those norms:

In the 1960s as today [2000], a substantial fraction of the population thought abortion evil, another substantial fraction thought it at worst a small wrong, and still another substantial fraction found itself torn. In the 1960s the first group had the law on its side; today the second group does. In each case the legal regime generated sympathetic cases for the losing side, cases that seemed to highlight the downside of the existing law—deaths from back-alley abortions then, borderline infanticides now. Those cases are sympathetic only because there is some constituency prone to feel sympathy; they cannot create opposition on their own. But given some base of opposition, the stories are bound to surface, and are bound to generate a reaction, because they deal with the sorts of heartrending events that would push fence-sitters in one direction or the other.¹⁴

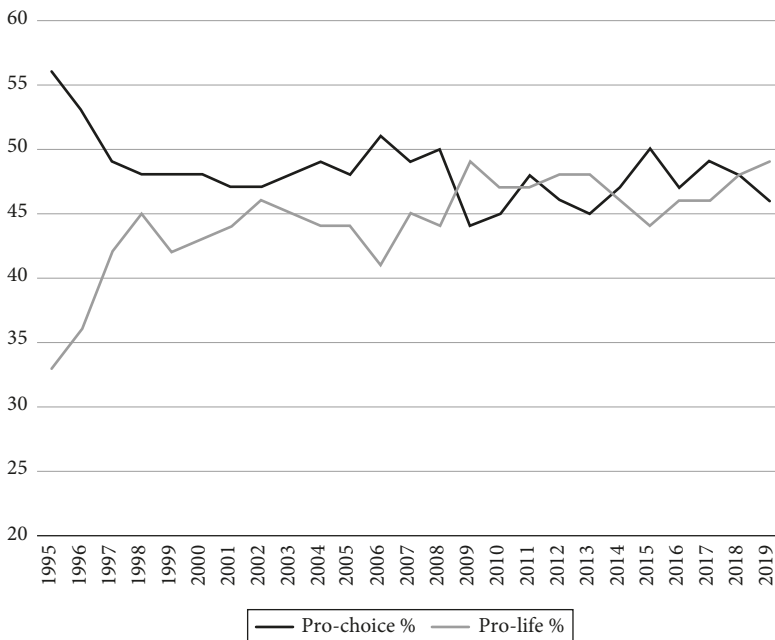
¹² See Schwab and Ostrom, “The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies,” pp. 209–11.

¹³ See Friedman, *Moral and Markets*, chap. 5.

¹⁴ Stuntz, “Self-Defeating Crimes,” p. 1888.

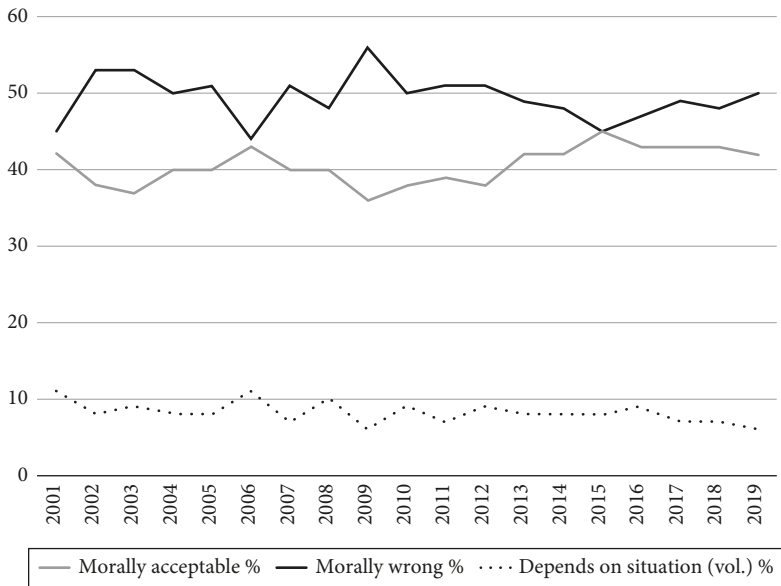
Displays III.1–III.3 provide a time series of American self-attributions of themselves as pro-life or pro-choice (III.1), judgments of whether abortion is morally wrong or permissible (III.2), and judgments of whether abortion should be legal (III.3).

The law, which has now been in place for half a century, has not been effective in pushing people's informal judgments to become more favorable to abortion or to the law. As Display III.1 indicates, the American public has never been more apt to describe itself as pro-life. More describe themselves as pro-life (49%) than pro-choice (46%), whereas in 1995, 56% described themselves as pro-choice and only 33% as pro-life. Attitudes about the morality of abortion have not shown as clear a trend, with a good deal of variance over the years (the third category of “depends” also complicates comparison), but it certainly has not been the case that the law has brought people around: more continue to believe it is morally wrong than permitted. Judgments of what the law ought to be show a relatively stable majority that it should be legal under some circumstances. Interestingly, however, considerably fewer believed in 2019 that it should be legal in all circumstances than in 1992 (25% vs. 43%), while more believed that it should be illegal in all circumstances (21% vs. 15%). Again, the law's effects seem to be somewhat self-defeating. Americans have gone from a 28% gap between those who believe abortion should always be legal and those who believe it should never be legal to a 4% gap.



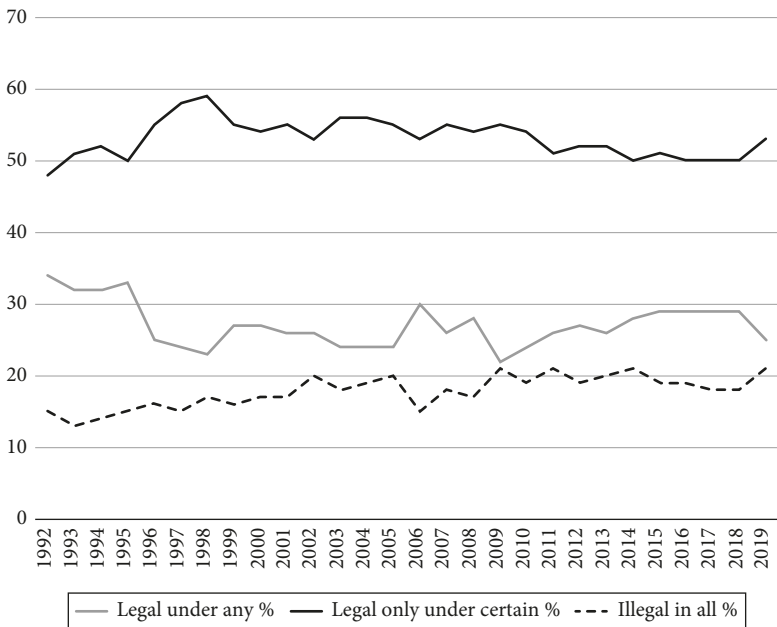
Display III.1. Changes in self-descriptions as pro-choice or pro-life.

“With respect to the abortion issue, would you consider yourself to be pro-choice or pro-life?” For all figures, when multiple polls were taken during the same year, the average is given. Some residue categories with very small number of respondents are omitted. Source: Gallup Poll <<https://news.gallup.com/poll/1576/abortion.aspx>>.



Display III.2. Changes in judgments of the morality of abortion.

“Regardless of whether or not you think it should be legal, for each one, please tell me whether you personally believe that in general it is morally acceptable or morally wrong. How about abortion?” The “depends” answers were voluntarily given by respondents. Source: Gallup Poll <<https://news.gallup.com/poll/1576/abortion.aspx>>.



Display III.3. Changes in judgments of the proper legality of abortion.

“Do you think abortions should be legal under any circumstances, legal only under certain circumstances, or illegal in all circumstances?” Source: Gallup Poll <<https://news.gallup.com/poll/1576/abortion.aspx>>.

The upshot is that the law sometimes strengthens the very norms and attitudes it sought to displace, while engendering skepticism about the law itself. Other examples abound.¹⁵ Even when the norm-opposing law is not straightforwardly self-defeating, evidence from social science indicates that laws running against moral norms and moral convictions are typically ineffective in changing people's behavior. Law-abidingness depends on the proper fit between moral convictions, norms, and legal rules. The well-known Chicago study measured individuals' (self-reported) law-abidingness with respect to a range of relatively mundane crimes—speeding, parking illegally, violating noise ordinances, littering, driving under the influence, and petty shoplifting—as well as their (self-reported) attitudes relating to the competing mechanisms of legal compliance posited by the legalistic and normative perspectives. In particular, the investigators measured individuals' beliefs about how likely they were to be punished for breaking particular laws (fear of punishment), the extent to which they thought peers would disapprove of them for breaking those laws (social norms), whether they believed the behaviors those laws prohibited were wrong (personal morality convictions), and whether they believed they had a general obligation to obey the law (legitimacy convictions).¹⁶ The study involved 1,575 residents of Chicago in the first wave, and 804 respondents (randomly selected from the first group) in the second; the purpose of the first wave was to measure the extent to which these different attitudinal variables explained differences in legal compliance between individuals, the purpose of the second was to measure the extent to which changes in attitudes explain changes in law-abidingness.¹⁷ On the basis of both waves of data, Tyler concludes that fear of punishment, social norms, and moral convictions all explain some variance in legal compliance, with moral convictions best explaining such variance, peer disapproval (or social norms) coming second, and fear of punishment last.¹⁸ Within the category of moral convictions, both attitudes toward the wrongness of particular actions and to the legitimacy of law in general appear to influence behavior, with the former playing a larger role.¹⁹ The Chicago study is only one of a group of studies that have come to essentially the same conclusions.²⁰

Again, the point is most definitely not that law is entirely ineffective in inducing norm and behavior change, or that it does not have a critical place in self-governance. However, the almost exclusive focus on law in most accounts of governance is deeply mistaken, and has done great damage in practical contexts. The law must work alongside the informal, self-organized system of rules and personal normative convictions. This conclusion is by no means based simply on my analysis of self-organized morality. As the 2017 World Bank Report on *Governance and the Law* concludes, the perception that formal legislation and institutions are sufficient for governance and change is clearly mistaken: the interaction of these institutions with informal norm

¹⁵ Such as Prohibition in America, which led to increased use of alcohol, as well as increased tolerance for organized crime and hostility to the police by some ethnic groups. See Stuntz, "Self-Defeating Crimes."

¹⁶ Tyler, *Why People Obey the Law*, chap. 4.

¹⁷ *Ibid.*, p. 8.

¹⁸ *Ibid.*, p. 45.

¹⁹ *Ibid.*, chap. 5.

²⁰ See Robinson, "Why Does the Criminal Law Care What the Layperson Thinks Is Just?"

networks and cultural beliefs is critical for governance.²¹ The Report thus defines governance “as the process through which state and nonstate actors interact to design and implement policies within a given set of formal and informal rules that shape and are shaped by power.”²²

22.3. The First Dimension of Governance: Pursuing Policy Goals

The manifest inadequacies of the Legislative Theory of Democratic Self-Governance led us to a far more plausible conception suggested by the World Bank’s analyses, which we might call the Process Theory of Democratic Self-Governance:

- (i) *Popular Control*: In a democratic polity, the people “achieve control over the state.”
- (ii) *Governance Process*: State policy is a process by which the democratic state interacts with non-state actors and informal norm networks to pursue its designated social goals/valued results.

Note that the “central controller” (aka the government) is now one actor in a system of interactions. It cannot shape or control the complex order, but seeks to secure certain goals or valued outcomes identified by the democratic public by coordinating and bargaining with a variety of non-state actors and networks. The World Bank identifies security, growth, and equity as the goals of governance.²³

As with the Bank, governance is commonly conceived as goal-centered: the governor identifies social goals or ends, which policy is intended to secure. The critical feature of this understanding of self-governance is “the policy state.”²⁴ governance is a collective pursuit of goals identified by the democratic state. Effective policies secure the democratic state’s goals (“I tell this one, ‘Go,’ and he goes; and that one, ‘Come,’ and he comes”).²⁵ This familiar conception of governance is straightforward under the Legislative Theory, where policies are simply directives to accomplish things—things to be done by the populace. However, on the far more plausible Process Theory, while the goals are set by the democratic state, it only has the ability to maneuver within an interactive, often strategic milieu.²⁶ It can issue directives, but whether these directives will be obeyed, and if so in what manner they will be obeyed (there are many ways to conform to rules; §18), is not something the state can determine.

It is important to stress that on this understanding of “self-governance,” many in the governance area will see the governor’s (the government’s) goals as exogenously imposed by it. Democratic theories such as Rousseau’s have understood democracy

²¹ 2017 World Bank Report on *Governance and the Law*, p. 52.

²² Ibid., p. 41. Compare Jacobs: “I am using self-government in its broadest sense, meaning both the informal and formal self-management of society.” *The Death and Life of Great American Cities*, p. 114.

²³ 2017 World Bank Report on *Governance and the Law*, p. 4.

²⁴ In nineteenth-century legal discourse, the “police” or “policy” state was contrasted to the *Rechtsstaat*—the legal or justice state. Hayek invokes the idea of the liberal *Rechtsstaat* against the “planning state” in *The Road to Serfdom*, p.117n.

²⁵ Elster, *Securities against Misrule*, p. 19.

²⁶ 2017 World Bank Report on *Governance and the Law*, chap. 2.

in terms of polities united by a common good, in which after a democratic vote the minority embraces the decision of the majority as “the general will”: the aim is for all to see the law as coming from all and applying to all.²⁷ However, for any theory of democracy employing an approximation to majority voting, its actual workings are bound to generate numerous cases (indeed the large majority) in which the government has declared that “we” have selected a goal that is not, as far as many reflective citizens are concerned, their goal. Important sections of the public are apt to reject many policies, that, say aim at equity or growth. To be sure, the governor will often seek to incentivize compliance, but there are, again, a variety of ways to respond to incentives, and it will be impossible to incentivize many people on many issues. In many cases, citizens will react strategically to the directions of the governor. In thinking about the possibilities for democratic self-governance, this type of reflexivity cannot be idealized away as “non-ideal” imperfect compliance.

22.4. The Second Dimension of Governance: Solving Strategic Dilemmas

Consider again Ismael’s notion of self-governance: “a centralized process that involves the sharing of information and the formation of an overall plan and deliberate coordination of joint activity . . . at least some of the information distributed throughout the systems is collected, synthesized, and used to fuel a decision procedure that plays a role in guiding the system’s behavior.” This core notion of “governance” sees the governor collecting information and attempting to guide the system—in a policy context, seeking to achieve certain system goals, often through a combination of directives, regulations, and incentives. As I have said, reflexivity will be a persistent problem for such goal-oriented governance, as citizens decide for themselves how they will respond to the governor’s moves. We must always remember that on the Process Theory of Democratic Self-Governance, the governor is one agent among many, not a true “controller.” Now this problem of reflexivity is far less serious when the governor is undertaking a different task: to assist subjects in resolving their strategic dilemmas. This, I shall argue, is a distinct dimension of system governance, with its own possibilities and pitfalls. Consider three categories of strategic dilemmas: *suboptimality traps*, *assurance problems*, and *coordination problems*.

Classic cases of suboptimality traps include dilemmas concerning the provision of public goods, club goods, and common resource problems. To briefly review the familiar, public goods are defined in terms of two characteristics. First, they are

²⁷ Jean-Jacques Rousseau, *The Social Contract*, esp. Books I and IV. Theories that employ the “Condorcet Jury Theorem” also can be interpreted as having this upshot. See further my “Does Democracy Reveal the Will of the People?” For two reasons I also put aside here theories of governance inspired by unanimity requirements such as James Buchanan’s: (i) I wish to examine something like real-world democratic policy, and these are not unanimity-inspired systems; (ii) in unanimity-inspired systems, governance is essentially reduced to solving strategic problems and specifying rules of the game, dimensions of governance that I shall consider in subsequent sections. Thus Buchanan’s governance system will be considered in due course. See Buchanan and Tullock, *The Calculus of Consent*, esp. chap. 7; Buchanan, “Taxation in Fiscal Exchange.” See also my *Order of Public Reason*, pp. 538–44.

characterized by *non-rival consumption*.²⁸ Consider clean air. If it is provided at all, it can be provided to Alf without taking any of it away from Betty. Once the good is there, consumers do not compete for it; everyone can freely use it without diminishing the amount left for others. Second, we cannot control the flow of benefits from public goods: they are *non-excludable*. If a public good is provided, it is provided for all to use. If we clean the air, everyone has clean air. We cannot exclude those who have not paid their share. A pure public good is one that perfectly meets these two conditions; a pure private good is one that does not meet either. Display III.4 identifies four classes of goods based on these two characteristics.

	Excludable	Non-excludable
Rival	Private Goods	Common Pool Resources
Non-rival	Club Goods (or “toll” goods)	Public Goods

Display III.4. Four classes of goods.

Private Goods: Private goods are the typical market good. If you go to the store to purchase milk, (1) the good is excludable—those who don’t pay for it do not get it; (2) it is rival. If I consume that carton of milk, you cannot. These do not generally present suboptimality traps. Individual action can produce optimality.

Club Goods: Some goods are non-rival within a group, but people can be excluded from the group (or “club”). Think of a movie theater (with good stadium-style seats, so all can easily see the movie). Within the theater, consumption is non-rival: if I “consume” the movie, there is no less of a movie for the person next to me to “consume.” One person’s consumption does not detract from that of others.

Common pool resources: Elinor Ostrom, who won the Nobel Prize for economics in 2009, studied common pool resources such as fisheries. With such resources, my use detracts from the resources available to you (if I take a fish out of the ocean, I have consumed it so you cannot have it, thus our consumption is rivalrous), but it is very hard to exclude anyone.

Public goods have both characteristics. A *pure* public good is one that perfectly meets these two conditions. In most cases, both conditions are not perfectly met, and this may matter a good deal for policy. Defense, law and order, regulation of air pollution, highways, ports, public works, and elementary education are usually cited as public goods, though economists have disputed the “publicness” of every item on this list.

²⁸ This feature has been characterized variously in terms of joint supply, joint consumption, consumption externality, and non-rival consumption. Some consider these essentially equivalent, but on some interpretations they diverge. See Head, *Public Goods and Public Welfare*, p. 228.

Even if everyone prefers having the public good (and paying for it) to not having it, each of us will receive it for free if enough others pay for it. Economists have long insisted that we each have an incentive to *free-ride*: each, hoping others pay, holds back from paying. Consequently, state provision of public goods can give everyone what they want: to get the good (and pay for it) rather than not getting the good at all. Here, the state can be seen as extracting us from a destructive Prisoner's Dilemma: the "I defect, others cooperate" (I get the good for free) strategy is (partially) eliminated. Reflexive strategizing is reduced.

In fact, experimental evidence indicates that in market-oriented societies, at least half the population do not initially seek to free-ride,²⁹ instead voluntarily contributing to public goods when others do (as we would expect from conditional cooperators). Yet because some do free-ride, cooperation tends to break down as conditional reciprocators withdraw their cooperation.³⁰ The cooperators prefer contribution to defection, but do not wish to be suckers (they are playing Stag Hunts). So the governor's role is best seen as providing *assurance* to cooperators that the smaller defecting population will not be allowed to free-ride. The governor is, then, is solving the strategic dilemma in two ways: it is providing assurance to the cooperators that others will not be allowed to free-ride, and it incentivizes the free-riders not to take up their preferred, but ultimately destructive, defect option, which is good for them too. The problem of reflectivity is especially mitigated for the cooperators (something like 50%–80% of the population), since the governor is assuring them of the outcome they seek. Only the part of the population who actually prefer the free-riding strategy are apt to devise reflexive responses that might undermine the governor's aims, and even they may see that the governor is helping them secure their aims.

Coordination problems also involve strategic dilemmas: there are many ways we might coordinate, but we need to select one. I have argued (§21) that moral communities can self-organize in solving impure coordination games: that there are multiple equilibria on which to coordinate does not require a governor to make a decision. As Hume stressed, informal social processes often lead to conventions, which select one of many possible equilibria as the social rule or practice.³¹ But again, this is not inevitable, and it may take time. A governor can be effective in selecting and stabilizing an equilibrium. It has the advantage of being able to act swiftly—and the great disadvantage of only guessing what people's preferences are, and so always being uncertain whether its actions represent the imposition of a policy goal (our first dimension) on many dissenting people, or whether it is truly helping a population solve its coordination problem. The world does not come with nice game matrixes to show everyone's payoffs; the governor is constantly acting with uncertainty about whether it is pursuing a controversial policy aim or solving a strategic dilemma of a population.

²⁹ Fischbacher, Gächter, and Fehr's study found that 50% were conditional cooperators, and about 30% free-riders. The remaining had more complex preferences. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." In other experiments, about 80% of the participants seem to be conditional cooperators. See Chaudhuri, "Sustaining Cooperation in Laboratory Public Goods Experiments."

³⁰ In addition to the sources in the previous note, see Ledyard, "Public Goods: A Survey of Experimental Research."

³¹ Hardin, *David Hume: Moral and Political Theorist*, chap. 4; Vanderschraaf, *Strategic Justice*, chap. 2.

22.5. The Third Dimension of Governance: The Rules of the Game

On the most familiar dimension of governance, the governor identifies preferred states and seeks to move society toward them. On the second, the governor seeks to solve strategic dilemmas to assist citizens securing the outcomes they all wish. On the third dimension, governance seeks to structure (some of) the rules of self-organization.

All organized activity by humans entails a structure to define the “way the game is played,” whether it is a sporting activity or the working of an economy. That structure is made up of institutions—formal rules, informal norms, and their enforcement characteristics.... How the game is actually played depends not only on the formal rules defining the incentive structure for the players and the strength of the informal norms but also on the effectiveness of enforcement of the rules. Changing the formal rules will alter the way the game is played but also, as anyone who has watched professional football knows, it frequently pays to evade the rules (and deliberately injure the quarterback of the opposing team). So it is with the performance characteristics of an economy. To understand performance we must explore in depth the way institutions “work,” looking at both the consequences of formal incentives and the frequently unanticipated results.³²

North reinforces our criticism of the Legislative Theory of Democratic Self-Governance: the state’s ability to legislate and coerce by no means shows it can dictate the actions of its citizens, or confidently predict the outcomes of its enactments. It is always engaged in strategic relations with the rest of society and its informal networks.

North is also correct that a governor is often critical in helping to construct this “scaffolding” of human interaction.³³ When laws conform to norms, or where there is a strong norm of legal obedience,³⁴ state action can be highly efficacious in shaping the rules of interactions. These not only constrain actions through prohibitions, but as the “rules of the game” metaphor suggests, they can empower individuals to do things that were previously undoable. Before there are established property rights, investment strategies and banking are not conceivable, just as without the rules of cricket there cannot be a “wide ball.”

Often the rules are instituted on essentially “deontic” grounds, or because the actions they prohibit are seen as intrinsically wrong, or the rules protect what is intrinsically valuable. “Most aspects of governance are valuable in and of themselves—that is, they have intrinsic value—in particular the notion of freedom.”³⁵ Aims such as the abolition of trafficking in women can generally be secured if compliance with the rule can be secured. Here the rule is not seeking to guide the system toward any goal—other than compliance with the rule itself. Yet the governance unit often has goals in instituting or revising the rules (or the set of rules that we call “institutions”), and in such cases the distinction is blurred between institutions as mere means to policy goals (our first dimension) and institutions as scaffolding allowing wide range of

³² North, *Understanding the Process of Economic Change*, p. 48.

³³ Ibid.

³⁴ Mackie, “Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience.”

³⁵ 2017 World Bank Report on *Governance and the Law*, p. 4.

outcomes (our third). After the 1968 “year of the pitcher,” major league baseball (the governor) decided games were too boring and defensive (a policy aim): it lowered the pitcher’s mound to decrease the advantage of pitchers (a basic rule of the game was changed), leading to higher scoring games. Nevertheless, there is an important distinction between the governor pursuing policy goals and setting up structures of interactions. At the extreme or pure case, a policy goal has a maximand: the policy, within the constraints, seeks to maximize some value (such as growth, longevity, welfare, clear air), and within cost constraints, adjustments that increase that value should always be pursued. With a policy goal, metrics make sense to see how well it is being secured. In contrast, in quintessential structural governance, the governor has some broad criteria for adequate functioning (“the rules shouldn’t make the game boring”), and then allows the game to be played, allowing for exciting and boring games.

22.6. The Space of Self-Governance

Our analysis yields a mapping of the space of self-governance. We have a general 3×3 space (three dimensions, three levels) to investigate. The general space is helpful in thinking of types/scope of self-governance, distinguishing our three dimensions of governance at the macro, meso, and micro levels of social life, and then distinguishing different problems to be explored in each space. As Display III.5 indicates, in this Part there are two sections discussing macro issues (§§23, 24), three focusing on the meso level (§§25–27), and one concerned with the micro (§28). The specific entries in each cell are what I take as important, though by no means exhaustive, problems

	<i>Macro</i>		<i>Meso</i>			<i>Micro</i>
	<i>Macro Control</i> (§23)	<i>Macro Structure</i> (§24)	<i>Strategic Dilemmas</i> (§25)	<i>Meso-level Goal Pursuit</i> (§26)	<i>Sectoral Goal-Governance</i> (§27)	<i>Micro Self-Governance</i> (§28)
<i>Goal-Policy</i>	Control/ guidance to secure overall system states			Like-minded communities pursuing common goals	Changing targeted variables: system- subsystem level	Targeting norm networks to secure goals
<i>Strategic</i>			Solving larger strategic problems through polycentric rule systems			Solving local strategic problems
<i>Rules/ structure</i>		Structuring the overall rules of the game				Reforming the rules of individual norm networks

Display III.5. A heuristic for analyzing the space of self-governance and a map of this Part.

of self-governance that I shall discuss. The general space could be further subdivided. Although such taxonomies are artificial and if taken too seriously can end up as straitjackets, they help us avoid the temptation of lapsing into sweeping claims about whether self-governance is or is not possible in complex systems. Theorizing about governance in complex systems is at the point where only more fine-grained analysis can show the way forward. I shall argue in this Part that Hayek was entirely correct that some categories of governance are impossible in complex systems, yet others are both possible and necessary. By regimenting discussion along the lines of this space, can we bring rigor to this vexed issue.

As I shall understand it, then, *sectoral governance* concerns some feature of the system or the subsystem that the governor seeks to manipulate (e.g., economic growth, health); *systemic governance* involves an attempt to manipulate the entire system as a whole (e.g., to secure a just and well-ordered society); *subsystem governance* involves an attempt to manipulate an entire subsystem (e.g., a regional development agency, the rules of a state constitution), while *micro governance* focuses on small, often face-to-face, networks. Of course these categories have fuzzy boundaries; my aim is not to insist on classificatory boundaries, but to avoid some overgeneralizations characteristic of governance discussions.

§23 Macro Control

23.1. Governance as Global Control

Our first category is the most straightforward: governance as controlling the overall direction of an entire social order. Humans wish to control their destiny,¹ and the most ambitious expression of this wish is to guide the social order along a path of overall improvement. This has been an enduring aim in political philosophy. Since Plato, philosophers have constructed utopias as guides to long-term social reform.² We will see presently that this has been not only a philosophical ambition: the development literature of much of the twentieth century manifested a similar project. Such is the grandest of all policy aspirations: self-governance leading to a new and better society. That is the concern of this section.

The following competencies would be required to secure, as Ismael puts it, an effective “overall plan and deliberate coordination of joint activity” for this grandest of all policy aims:

- (i) The central controller must be able to monitor and aggregate information from diverse sources at time t_1 to arrive at a significantly better than random judgment of the overall system state S_1 at t_1 ;

¹ North, *Understanding the Process of Economic Change*, p. 7.

² I have analyzed this ambition in some depth in *Tyranny of the Ideal*, chaps. 1–3.

- (ii) The central controller must then decide on the overall goal for a new system state G (an “overall plan and deliberate coordination of joint activity”) that it seeks to achieve, an intervention i_1 , that has a significantly better than random chance of producing (or moving closer to)³ a system state S_2 that achieves (or is closer to) G at t_2 .
- (iii) The central controller must be able to monitor and aggregate information from diverse sources at time t_2 to arrive at a significantly better than random judgment of the system’s state S_2 to determine whether it is indeed closer than S_1 to the desired G condition. (At this point the central controller may decide to reaggregate information and revise the desired goal to G^* .)
- (iv) If it is judged “not closer,” the central controller must be able to monitor and aggregate information from diverse sources at time t_2 to arrive at a significantly better than random judgment as to an intervention i_2 that has a significantly better than random chance of producing S_3 that achieves (or moves closer to G/G^*) at t_3 .

These specific competencies manifest two broader competencies or functions: “cognitive” and “manipulative.”⁴ “As Beinhocker puts it, *‘If I perceive state A (cognitive function) and take action X (manipulative function) then state B will result, bringing me closer to (or farther from) my goal G.’*”⁵ If a centralized decision-making procedure can do this it can pursue systemic, global, policy goals, guiding the system in a preferred direction.

It is easy to suppose that this entails a command-and-control model, where the self-governor sets a goal and then draws on social scientific and policy experts to devise rules and institutions to control the social order so as to secure it. This, though, seems the least plausible version of goal-oriented self-governance: it would require the central controller not only to have comprehensive knowledge of current and future social contingencies and the causal levers needed to alter them to bring about the goals, but also to have a well-worked-out and clear notion of the system-level goal and the outset of the policy design process. I know of no complexity theorist who thinks that the cognitive and manipulative resources posited by (i)–(iv) are to be had in complex societies. The amount of up-front information about both the goal and system behavior in response to interventions would be enormous: the governor would have to have a clear idea of the goal and would have to have identified the appropriate cognitive and manipulative levers at the outset. Much more plausible are theories of governance that analyze it as a recursive process “of provisional goal-setting and revision based on learning from the comparison of alternative approaches to advancing them in different contexts.” The concept of a recursive function is borrowed from mathematics “whereby the output from one application of a procedure or sequence of operations becomes the input for the next, so that iteration of the same process produces changing results.”⁶ Here not only are the effective means to secure the goal constantly

³ On metrics of “moving closer to a desired state,” see my *Tyranny of the Ideal*, pp. 4–61, 251–9.

⁴ Soros, “Fallibility, Reflexivity, and the Human Uncertainty Principle.”

⁵ Beinhocker, “Reflexivity, Complexity, and the Nature of Social Science,” p. 332. Emphasis in original.

⁶ Sabel and Zeitlin, “Experimentalist Governance,” p. 169.

revised in the face of experience, but the goals themselves are refined as governors reflect on efforts to secure them.⁷

This broadly Deweyian-inspired notion of policy as a learning process is at the heart of Knight and Johnson's recent important analysis of democratic self-governance. They see centralized democratic decision-making as aggregating information dispersed throughout the system to evaluate the system's functioning and employing the information to "experimentally" reform it in socially desirable ways (see further §28.1). The tasks of centralized democratic decision-making are "(1) to coordinate effective institutional experimentation, (2) to monitor and assess effective institutional performance for the range of institutions available in any society, (3) to monitor and assess its own ongoing performance."⁸ In addition to the cognitive functions of (2) and (3), the democratic self-governor must have a sufficiently strong manipulative ability to make experimental adjustments that have a significant likelihood of producing reform in the desired direction. The stress on institutional "experimentation" is fundamental to their essentially pragmatic approach: democratic self-governance does not simply select a path to be followed, but is an ongoing process of monitoring, evaluating, and "experimenting" to discover better institutional arrangements that yield better overall social results, as well as a better understanding of the macro goals themselves. Such centralized reflexive capability is held to give democratic decision-making a priority over all forms of decentralized self-organization.⁹ Our question here is whether this experimental approach could allow a society to, as North put it, "control their destiny" by guiding the overall direction of their society. Later we shall consider more modest applications of it.

23.2. Institutional Complexity

A great strength of Knight and Johnson's analysis is their stress on institutional evaluation. In a complex modern society, the critical channels of self-governance are not edicts or directives, but institutions. The role of recursive democracy is not to instruct citizens what to do to secure social goals, but to monitor institutional performance, including the performance of the democracy itself, and to alter institutions to better conform to system goals.¹⁰ Let us start with an idealized distinction between formal and informal institutions, where the former are largely the creations of the legal-administrative authority of the state, while the latter are the upshot of self-organized processes.¹¹ In this subsection I shall focus only on the formal institutions, and make

⁷ See Levy and Peart's criticism of the "linear" model of policy. *Escape from Democracy*, p. 8.

⁸ Knight and Johnson, *The Priority of Democracy*, p. 169.

⁹ They also hold that self-organized systems cannot ensure society-wide coordination (*ibid.*, pp. 105–6), and thus centralized self-governance is required to ensure the normative acceptability of the overall system. Cf. Wilson's view in §17.3.

¹⁰ It is thus both recursive and reflexive: it recursively takes one period's evaluative judgments and decisions as an input into the next period's, but at the same time reflexively monitors its own performance. Of course the same processes can be applied at the subsystem level.

¹¹ I am broadly following North, *Understanding the Process of Economic Change*, pp. 42ff.

two Herculean assumptions: (i) that the governor has manipulative capabilities to actually change the institutions in the direction it wishes,¹² and (ii) that individuals comply with the requirements of the formal institutions.

If the functioning of each of the main institutions are decomposable (§17.3)—their functioning is not intimately tied to the others (or, as I shall now say, they are *modular*)—the recursive/reflexive process would be an intelligent approach to systematic self-governance. To say a formal institution F_i is *modular* in system S is to say that given a total set of institutions $F (F_1 \dots F_n)$, a reform of F_i has no effect on the functioning of the other formal institutions in the set, and that changes in the other institutions do not affect F_i 's functioning. If such modularity holds, it is not too much of a stretch to say that the governor can have knowledge that a change in F_i to F_i^* can alter its functioning so as to better secure social goal G . Even if F_i were *quasi-modular*, this could be so. F_i is quasi-modular if (i) a change in F_i affects the functioning of only a small number of other formal institutions in the set, or a change in F_i only has marginal effects on the functioning of a number of other formal institutions in F , and (ii) changes in other elements of F do not have profound effects on F_i .

As we saw in section 17, modularity (which is implied by decomposability) seems an upshot of successful macro selection (or, alternatively, is a prerequisite for such evolution). If the system was selected as a whole, it certainly appears that it must be quasi-modular for the reasons that Simon advances: if the system becomes too complex, a change in one feature can radically change the whole, and so it cannot climb an evolutionary gradient. But we have seen that self-organized complex systems are constantly evolving greater complexity: as we reform our institutions to work together to solve ever-more difficult problems, they become ever-more interconnected and so complex. As Miller observes, for example, “[w]e have unknowingly created a complex adaptive financial system that we do not understand and cannot control. At each stage of its creation, we have accrued additional complexity in the name of added benefits.”¹³ Charles F. Sabel and Jonathan Zeitlin also point to a

secular rise in environmental volatility and complexity over the past few decades. Some of this can be linked directly to globalization, such as the problems of managing transborder common-pool resources like water or of ensuring the safety of imported food and other products as they move through transnational supply chains. In other cases, the transnational connection is only part of the story, as with the accelerating pace of technological innovation ... or the diversification of household and family structures, employment patterns, and populations.¹⁴

The attempt to secure goals through a set of interconnected institutions leads us back to our NK optimization problem—we are optimizing over N institutions with K interdependencies among them (§17.2).¹⁵ When $K = 0$, that is, when there are no

¹² This implies the further Herculean assumption that the state is a unitary actor. See Platteau, *Institutions, Social Norms and Economic Development*, p. 177.

¹³ Miller, *A Crude Look at the Whole*, p. 62. Cf. Neil Johnson, *Simply Complexity*, chap. 6.

¹⁴ Sabel and Zeitlin, “Experimental Governance,” p. 174.

¹⁵ Kauffman, *The Origins of Order*, especially chap. 2.

interdependencies (there is perfect modularity) in the performance of the institutions, advancing toward the social goal is straightforward (given our strong assumptions!): we can adjust each one, *seriatim*, confident that today's modifications of F_2 will not undo yesterday's adjustment of F_1 . But when most institutions affect most others, it has been proven that the optimization problem is exceedingly complex: the possibility space abounds in poor combinations. Moreover, incremental recursive adjustments will almost surely lead to "dead-end" or "poor local optima" where (i) the social goal is not well secured and (ii) all incremental adjustments decrease its achievement. Incrementalism is by no means a route to constant improvement in such systems.¹⁶

High *NK* systems, we have seen, approach the chaotic. A rather overused example is Lorenz's famous "Butterfly Effect:" the flap of a butterfly's wing in Brazil could cause a tornado in Texas.¹⁷ But anyone living through 2020 will have a much more drastic case: as far as we can tell, the workings of a Wuhan wet market led to millions of deaths, a global economic catastrophe, a change in the life of almost everyone on the planet, crises in higher education, and political conflicts within and between countries. A person selling a snake or a bat in Wuhan apparently led to a crisis not simply in health, but in economics, education, politics, and social life in general. The demonstration that the world is an unpredictable complex system is obvious for all to see, yet even today many experts continue to think they have access to the cognitive and manipulative functions to predict and control the system.

The gales of uncreative destruction of 2020 highlight two critical features of the contemporary Open Society. In his 2004 *Fooled by Randomness* and his 2007 *Black Swan*, Nassim Nicholas Taleb developed the idea of a Black Swan: a very rare, entirely unexpected, random event with large systemic impact.¹⁸ We have now witnessed two destructive "Black Swan-like" events in the last twelve years: in 2008 and 2020.¹⁹ As our system becomes ever-more complex and interconnected—and so decomposability continually decreases—such events become more common. As Taleb argues, *post hoc* we are convinced we should have seen this one coming, but the next one takes us equally by surprise. Like the French strategists who built the Maginot line, we are continually preparing to battle the last thing that took us by surprise. As the EU chief economist said in 2020, "Our tools were designed for a different crisis."²⁰

But, secondly, while a "Black Swan-like" event should jolt us awake to the ever-increasing connectivity and macro-unpredictability of our society, the underlying

¹⁶ See also Jervis, "Complexity and the Analysis of Political and Social Life," p. 574. I return to the *NK* model in §26.

¹⁷ See Lorenz, *The Essence of Chaos*. See also Smith, *Explaining Chaos*. Complexity is sometimes described as falling between "order and chaos," as for example in Mitchell, *Complexity*. On the other hand, many analyses of complexity include elements often ascribed to chaotic systems, such as sensitivity of the system to minute differences in the states of its members.

¹⁸ Taleb, *The Black Swan*; Taleb, *Fooled by Randomness*, p. 61.

¹⁹ We need not focus too much on totally "unexpected"—even if we know that some event can happen, if we cannot give any probability as to when, where, how, or how far-reaching it will be, we cannot adequately prepare for it. There are countless things that we know *can* happen and probably eventually will: the question is, of the countless things that we know can happen, which are to be prepared for now—and how.

²⁰ "Brussels Debates Options for Joint Rescue Package," *Financial Times*, March 23, 2020.

connectivities are always there, and defeat our everyday efforts to steer our system. It was not that in 2020 the world suddenly was linked with a multiplicity of institutions working to produce unpredicted outcomes—it was that they all contributed to an obvious catastrophic outcome. All the time these effects are flowing back and forth, combining and recombining, to defeat all efforts to move a society toward systemic social goals.

23.3. A Micro Example

Thus far I have not considered the complexities posed by citizen reflexivity. As I stressed in section 22.3, in politics with majoritarian-inclining political decision-making, many citizens will perceive the declared social goals as exogenous—as imposed on them by an external authority. In these cases, the reflexivity of citizen responses will be crucial: citizens will decide their best responses to institutional requirements, and these cannot be controlled by governments. It is always important to remember that such reflexivity is not simply a matter of whether they will obey but, if they obey, how they will do so.

To make vivid how much it matters how citizens respond to institutional requirements and how difficult this is to predict, consider a micro-example that looks ridiculously easy: mandatory seat belt laws that aim to decrease road deaths. Do they? The debate about the efficacy of seat belts began with Sam Peltzman's 1975 paper in which he argued that new cars subject to stricter safety regulations accounted for a significantly larger share of accidents than cars of similar age before safety regulations were implemented. Peltzman insisted that there was no evidence to show that seat belts saved lives.²¹ The crux of Peltzman's analysis is the hypothesis of offsetting behavior: as cars become safer, drivers adjust their behavior and drive less safely, thus offsetting the effects of the increased safety provided by seat belts. This is classic reflexivity: having created a safer environment, Peltzman argued, drivers adjusted their behavior to it. His offsetting-behavior hypothesis is disputed in the literature: David J. Houston, Lilliard Richardson, Jr., and Grant W. Neely suggest that there is little evidence for it;²² A. C. Harvey and J. Durbin, as well as David L. Ryan and Guy A. Bridgeman, report contradictory results;²³ while Christopher Garbacz's data provides strong support for it.²⁴ Steven Rhoads is confident that offsetting behavior occurs to some extent, though he acknowledges that non-economists have trouble accepting the idea. Rhoads reports that in Sweden, for example, it was found that drivers with studded snow tires drive faster than drivers with regular tires when the roads are icy, but not when they are dry.²⁵

²¹ Peltzman, "The Effects of Automobile Safety Regulations."

²² Houston, Richardson, and Neely, "Legislating Traffic Safety."

²³ Harvey and Durbin, "The Effects of Seat Belt Legislation on British Road Deaths"; Ryan and Bridgeman, "Judging the Roles of Legislation, Education, and Offsetting Behaviour in Seat Belt Use."

²⁴ Garbacz, "Estimating Seat Belt Effectiveness with Seat Belt Usage Data from the Centers for Disease Control"; Garbacz, "Do Front-Seat Belt Laws Put Rear-Seat Passengers at Risk?"

²⁵ Rhoads, *The Economist's View of the World*, p. 57.

If reflexive, offsetting behavior does exist, the question in evaluating whether mandatory seat belt legislation saves lives turns on whether the gains through greater probability of surviving a crash are offset by more crashes, or crashes of greater severity, or other new fatalities resulting from riskier behavior. Again, the evidence is contradictory. Garbacz finds that increased fatalities by rear-seat passengers more than outweigh the gains of greater safety of front-seat passengers,²⁶ while Harvey and Durbin find some increased risk for rear-seat passengers, pedestrians, and cyclists, but argue that it is not significant, and certainly does not outweigh the gains to front-seat belt wearers.²⁷ Although most studies have concluded that the gains outweigh the costs,²⁸ it is noteworthy that the reported gains are almost always much more modest than theoretical estimates (usually involving test crashes of dummies) predicted, providing support for the offsetting-behavior hypothesis.²⁹ There is also a wide, though not a complete, consensus that mandatory seat belt legislation has made life riskier for pedestrians³⁰ and cyclists as drivers reflexively adjusted their driving behavior.³¹

A recent debate has also arisen about the effectiveness of two types of “mandatory” seat belt legislation: “primary” regulations that allow the police to stop motorists simply on the grounds that they are not wearing their belts, and “secondary” regulations that only allow ticketing if a motorist is stopped for another infraction. Cohen and Einav and Houston and Richardson found that primary legislation can considerably enhance effectiveness.³² The issue seemed settled until 2017, when Harper and Strumpf found no evidence that changing from secondary to primary enforcement decreased road deaths.³³ Following Farmer and Lund,³⁴ they suggest that improvement in vehicle design and road improvements may have sufficiently changed the environment such that earlier dynamics no longer hold.

Perhaps to a philosopher this is more than a little tedious. In the end—after several decades—we seem to know there is a positive effect! But that is not the point: what looked like an absolute no-brainer simple regulation with a very clear goal has been excruciatingly difficult to evaluate—both concerning the effects of the policy and the

²⁶ Garbacz, “Do Front-Seat Belt Laws Put Rear-Seat Passengers at Risk?”

²⁷ Harvey and Durbin, “The Effects of Seat Belt Legislation on British Road Deaths.”

²⁸ Cohen and Einav, “The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities”; Rhoads, *The Economist's View of the World*, p. 239 n. 65.

²⁹ Irwin, “Technical Expertise and Risk Conflict”; Ryan and Bridgeman, “Judging the Roles of Legislation, Education, and Offsetting Behaviour in Seat Belt Use,” p. 32. For an exception, see Bhattacharyya and Layton, “Effectiveness of Seat Belt Legislation on the Queensland Road Toll.” Cohen and Einav do not find significant support for the offsetting behavior hypothesis. “The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities.”

³⁰ Rhoads, *The Economist's View of the World*, p. 239.

³¹ The Cyclists' Touring Club and the British Cycling Federation opposed renewal of mandatory seat belt legislation in the United Kingdom in 1986. See Irwin, “Technical Expertise and Risk Conflict,” p. 359. See also Jervis, “Complexity and the Analysis of Political and Social Life,” pp. 58–81.

³² Cohen and Einav, “The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities”; Houston and Richardson, “Traffic Safety and the Switch to a Primary Seat Belt Law”; Houston and Richardson, “Reducing Traffic Fatalities in the American States by Upgrading Seat Belt Use Laws to Primary Enforcement.”

³³ Harper and Strumpf, “Primary Enforcement of Mandatory Seat Belt Laws and Motor Vehicle Crash Deaths.” <https://pubmed.ncbi.nlm.nih.gov/28336356/>

³⁴ Farmer and Lund, “The Effects of Vehicle Redesign on the Risk of Driver Death.”

impact of reflexivity.³⁵ Reflect now on the scope for reflexivity in *steering a whole society to secure a comprehensive social goal*. The problems of reflexivity that seem to arise in seat belt laws pale into utter insignificance when we contemplate the project of steering an entire society along the governor's preferred path.

23.4. A Macro Example

We have considered the theoretical basis for the impossibility of effective systemic policy, and a micro example to illustrate some of the difficulties in anticipating reflexive responses. Yet, it might be wondered whether such systemic, truly global policy goals have actually been entertained. Political philosophers certainly contemplate them. Rawls believed that a "a long-term goal of political endeavor" should be to realize "the social world" with "the features of a realistic Utopia."³⁶ But systemic transformation has by no means been an exclusive ambition of philosophers: in the mid-twentieth century, not only socialists but a wide variety of political programs, from progressives to corporatists, thought a conscious centralized control of economy and society was both possible and necessary.³⁷ The "confusion, uncertainty and conflict" of "laissez-faire" was to be replaced by intelligence and experimental control, promoting the flowering of human capacities.³⁸ In the words of Keynes, the aim was to transition "from economic anarchy" to "a regime which deliberately aims at controlling and directing economic forces in the interests of social justice and social stability."³⁹

However, a far more important example of the pursuit of a systemic, overall social goal is the sustained and large-scale effort in the last sixty years to promote "development." Unlike the preceding cases, this was not simply a vision of ambitious theoreticians about securing social control, but a worldwide effort involving countless billions of dollars, often guided by the best economic and other experts of their time. Early theories of development, such as Rostow's *The Stages of Economic Growth*, depicted development as a series of social stages: traditional society; preconditions for takeoff; takeoff; the drive to maturity; the age of high mass consumption.⁴⁰ Although these stages were not strictly invariant, the supposition was that there was a natural course of progress in societies, and that "development" was a stage with common features achieved through a normal trajectory. While this stage theory has been abandoned, the very categorization of "developed" and "developing" societies echoes this supposition. As development efforts became more sophisticated, it was realized that, while there is nothing so neat as five stages of development, certain institutions were characteristic of all developed countries, such as stable property rights, the rule of

³⁵ This is just one of numerous cases that could have been presented. For a litany of such examples, see Tenner, *Why Things Bite Back*.

³⁶ Rawls, *The Law of Peoples*, p. 138. To this he adds, "[the] idea of realistic Utopia is importantly institutional." I have tried to show the futility of this aim in *The Tyranny of the Ideal*.

³⁷ See my "Private and Public Interests in Liberal Political Economy, Old and New," pp. 201–1.

³⁸ Dewey, *Liberalism and Social Action*, pp. 92–3.

³⁹ Keynes, "Am I a Liberal?," p. 305. See more generally Graham, *Toward a Planned Society*.

⁴⁰ Rostow, *The Stages of Economic Growth*, esp. chap. 1. Others depicted social development in terms of moral developmental stages. See Rosenberg, Ward, and Chilton, *Political Reasoning and Cognition*, chap. 7.

law, and mass education. As development thought proceeded, the result was an “ever-lengthening list of factors and conditions, of obstacles and prerequisites.”⁴¹ Thus, securing the goal of development required *institutional transplanting*: taking key institutions of developed societies and building them in the “less developed” or “developing” societies.

For decades, academics as well as practitioners have acknowledged the importance of *institutions*—organizations and rules—to development. Countries that are more secure, prosperous and equitable tend to rank higher on the existing indicators that emphasize certain institutional forms. The pattern has created a perception that certain types of institutions unambiguously determined higher levels of development, and it has led well-intentioned policy makers and development agencies to promote institutional standards—often referred to as *institutional transplants*.⁴²

As the World Bank Report goes on to note, transplanting has been a notoriously hit-and-miss affair: sometimes the institutional transplant took hold and worked as anticipated, while other times it failed and/or had serious detrimental effects. The development literature is filled with interventions that failed or backfired. The 2017 Bank Report argues that a complex relation between institutional forms and “underlying determinants” of a social order interact to determine whether development is achieved.⁴³ This, of course, requires the norms and beliefs to be in play that support the institution. Indeed, as the World Bank has stressed, institutions have a “schematizing role”; “they are not just rules, but also a way of seeing. Institutions shape the categories and concepts that people accept as natural and use to interpret the world.”⁴⁴ To truly “transplant” an institution would be to transplant shared mental models and schemas.

Thus far I have largely refrained from invoking the idea of an “emergent” property in the analysis of complexity, but “development” looks to be a quintessential instance. Emergent properties are sometimes distinguished from mere “resultant” properties on the grounds that, while a resultant property is the expected consequence of an underlying set of properties, emergent properties are very often novel and unexpected.⁴⁵ In perhaps the earliest analysis of such systems, John Stuart Mill considered a system, say, S , composed of elements (e.g., rules) $\{r_1 \dots r_n\}$ and an overall resulting order O .⁴⁶ Mill proposes three features of property O :

- (1) O is not the sum of $\{r_1 \dots r_n\}$;
- (2) O is of an entirely different character than $\{r_1 \dots r_n\}$;

⁴¹ 2017 World Bank Report on *Governance and the Law*, p. 40, quoting Albert O. Hirschman, *The Strategy of Economic Development*, who noted this in 1958.

⁴² 2017 World Bank Report on *Governance and the Law*, p. 52.

⁴³ *Ibid.*, p. 40.

⁴⁴ World Development Report 2015, *Mind, Society, and Behavior*, chap. 3. See also North, *Understanding the Process of Economic Change*, chaps. 3 and 4.

⁴⁵ See Waldrop, *Complexity*, pp. 81–3; Mitchell, *Complexity*; Page, *Diversity and Complexity*; Holland, *Complexity*, chap. 6. For application to policy, see Cairney, “Complexity Theory in Political Science and Public Policy.”

⁴⁶ Mill, *A System of Logic*, pp. 370–3; 438–40. For a helpful analysis, see Auyang, *Foundations of Complex-Systems Theories in Economics, Evolutionary Biology and Statistical Physics*, section 22. Hayek recognized such emergent properties. “The Theory of Complex Phenomena,” p. 269 and §10.4 in this volume.

- (3) O cannot not be predicted or deduced from the behavior of the members of $\{r_1 \dots r_n\}$ considered independently (i.e., apart from their interactions in S).

An emergent property O can be realized by a multitude of different micro-system states. Hence with development: we can judge whether a society is developed or “developing,” but two different developed societies may have very different micro bases. Devising a blueprint for development based on society S_1 will seek to identify some set of micro foundations, but we shall find that seldom is any particular set of micro features necessary to development. And if one, such as secure property rights, is common to all, the other micro foundations on which it rests are apt to be different in various systems. Thus, the inadequacy of transplanting some micro foundations to other orders: it is most unlikely that the emergent property can be thus induced.

It should be noted that the documented failure of institutional transplants producing a sought-after social emergent property casts grave doubts on macro theories of social evolution that stress diffusion via some groups copying the institutions and rules of more successful groups (§11.4). Anthropologists have documented such institutional transplants in small-scale cultures,⁴⁷ but it is fundamentally inconsistent with high complexity. As we have seen, copying rule R from a successful society is most unlikely to replicate the value of R in the copied-into society.⁴⁸ Consequently, for cultural evolution via inter-group copying of successful variants to work, culture cannot be a “tightly structured whole”—it cannot be too complex.⁴⁹ Indeed, it must be highly modular. This severely constrains the available selection mechanisms for the cultural group-selection thesis: conflict and immigration seem to have to bear the task of selection in complex societies.

We must appreciate the difference between Alf copying a cultural innovation x from Betty, and culture ALPHA seeking to copy institution X from culture BETA. When Alf copies a successful innovation of Betty’s searches, he seeks to learn what she had done and whether it will work for him. Sometimes he will fail, since he will not have correctly copied x , or does not have the background skills to use it. But in his society there will be many attempts to copy Betty’s x . If it cannot be replicated by many, it will die out; if it can and proves useful to others, it will grow. Social learning succeeds by trials, with many failures. But unless we have a very large set of cultures and learning possibilities, an attempt by culture ALPHA to imitate a cultural innovation of BETA is likely to result in our tale of world development: as cultures become complex, it becomes excruciatingly difficult for ALPHA to pick out from BETA an institutional package that has evolved in BETA that will work well in ALPHA. Numerous development programs devised by the best and brightest have been thwarted by complexity. As we shall see later in this Part, this does not mean that development cannot be promoted: micro experiments in ALPHA can help ALPHA find ways to revise its own norms and institutions to yield improved outcomes. The aim of the friend of

⁴⁷ See, for example, Henrich, *The Secret of Our Success*, pp. 173ff. See also Mesoudi, *Cultural Evolution*, pp. 68ff.

⁴⁸ Hayek recognizes this; *Rules and Order*, pp. 62–3.

⁴⁹ Richerson and Boyd, *Not by Genes Alone*, pp. 91–3. Cf. Boyd and Richerson, “Simple Models of Complex Phenomena.”

development should not be to argue against complexity, but to understand it so as to devise successful strategies.

§24 Macro Structure

24.1. The Systemic Framework

I have been highly skeptical of the ability of complex societies to successfully pursue valued overall systemic states, such as a well-ordered just society or development. However, we have seen that “governance” need not mean “control” or “goal pursuit.” Much more promising is our next category—system-level effective governance focused on specifying the “rules of the game,” which structure self-organization. As Hayek stressed—perhaps somewhat paradoxically—the rules and institutions of a complex society facilitate self-organization by removing some actions from the permissible. “The richer the artifactual structure, the wider the range of routine decisions that can be made,”¹ and so the more individuals can use their resources to explore the adjacent possible and to reflexively adjust to the actions of others. Although I have stressed that the rules themselves can be self-organized (§20), in many cases—especially in larger, systemic, contexts—a system governor can be highly effective in helping to construct this “artifactual structure.” As always, we must remember that while “formal institutions can be changed by fiat, informal institutions evolve in ways that are still far from completely understood and therefore are not typically amenable to deliberate human manipulation.”² Without coordination with the informal institutions, the governor’s “fiats” are apt to be ineffective (§22.2): as the development community has learned, there is a huge gap between institutional fiats and an effective artifactual structure. Formal-informal cooperative governance is typically required, but that does not obviate the need for a governor to play its part in developing a structure for self-organization.³

Because Hayek insisted that our ability to control and predict complex systems is so limited, he is often depicted as simply a skeptic about social knowledge.⁴ But this “skepticism” is based on a theory of complex systems. Complexity science does not simply tell us what we do not know and cannot do, but it tells us what we can know and can do. There is, of course, disagreement within complexity science as to what we know and what can be accomplished; it is an emerging field and we should not expect consensus. Some, for example, stress that because complex systems are nonlinear, they have “tipping points” in which the system can quickly change states: relatively

¹ North, *Understanding the Process of Economic Change*, p. 36.

² *Ibid.*, p. 50.

³ “[W]hat needs to evolve are not only the suitable practices and rules themselves, but also the public agencies (among which the centralized state is an essential yet not exclusive element).” Platteau, *Institutions, Norms and Economic Development*, p. xix.

⁴ For a sophisticated expression of this view, see Scheall, *F. A. Hayek and the Epistemology of Politics*.

minor actions can have very large effects in the system.⁵ Nevertheless, I believe there are at least three areas in which complexity analysis gives us some insights into governance qua implementing the rules of the game.

24.2. The Complex Society Moral Package

We have already considered in some depth the Complex Society Moral Package, and the way its bundle of rules facilitates the self-organization of a complex society. To know a system is complex is to know something about the ways it depends on certain sorts of rules. Natural liberty, property rules, harm rules—all are critical to complex self-organization (§18). In some cases, the rules of the Complex Society Moral Package can be self-organized, but often a coordination of moral and formal rules is required, especially concerning the overall framework for large-scale social life. It is the obvious importance of formal rules in many of these contexts that has led so many to fall into the error of confusing legislation and governance (§22.2); that legislation is often necessary certainly does not mean it is sufficient or decisive. And it is not even always necessary. Platteau finds that in some development contexts the “spontaneous individualization of land rights, unassisted by any process of titling or registration at the state level, can be extremely effective in activating the land market even when land sales do actually violate the law.”⁶ As always, the relation of law to governance is a matter of system dynamics: formal institutions are often critical, but are neither necessary nor sufficient for an effective implementation of the rules of the game.

24.3. Resilience

I have stressed that the development of autocatalytic diversity and complexity produces Schumpeter’s “gales of destruction” (often, though not always, creative; §17.2). A complex system is resilient when, in the face of shocks that destroy existing cooperative interactions, self-organizing searches discover new cooperative arrangements. *Resilience is simply self-organization at work*. Systems are highly resilient when the framework supports self-organization which, critically, requires what Taleb calls “optionality.”⁷ Natural liberty is, of course, critical to optionality, but it is not sufficient. Unemployment insurance and other aspects of a social safety net are required if, in the face of serious disruptions, individuals are to have live options to readjust their activities. While designing a framework for resilience is not simply a shot in the dark, it is always a tentative enterprise. The German *Kurzarbeit* scheme, which allows workers to be switched to part-time employment during a crisis with supplemental

⁵ For public policy relevant discussions, see Cairney, “Complexity Theory in Political Science and Public Policy,” p. 351; Colander and Krupers, *Complexity and the Art of Public Policy*, p. 54. For modeling, see Miller and Page, *Complex Adaptive Systems*, 143–6. For an optimistic appraisal of some engineering of complex systems, see Johnson, *Simply Complexity*, chap. 4.

⁶ Platteau, *Institutions, Norms and Economic Development*, p. 159. See also Libecap, *Contracting for Property Rights*.

⁷ Taleb, *Antifragile*, Book IV.

income from a fund, is said to have strongly contributed to German resilience in the 2008 crisis. Human capital is retained, allowing firms to re-establish themselves more quickly.⁸ Yet, in the 2020 crisis, the worry emerged that it was thwarting the system from reorganizing by freezing human capital in its pre-2020 uses.⁹ The great difficulty with resilience design is to minimize the tendency of the system to experience cataclysmic cascades¹⁰ while allowing people to effectively respond to the inevitable disruptions. In the midst of a severe crisis, governors tend only to care about the short term: if enough regulations and resources are thrown at the system, hopefully some will stick and work. But like the use of the military in a crisis, without an exit strategy the emergency measures from one crisis provide the rigidities aggravating the next. (Think again of the Maginot Line.)

A resilience-friendly framework does not seek to control the social order to prevent or smooth out disruptions but, in the motto of the British Idealists, to “hinder the hindrances” to self-organization.¹¹ The aim is not to guide self-organization, but to remove impediments to people taking up options and effectively adjusting their behavior. Prosaically, the aim is to ensure that the game continues in the face of unexpected events, not to guide the course of the game. A resilience-friendly framework will not seek to control the order to contain the gales. Efforts at control to reduce susceptibility to the constant shocks and surprises in complex societies tend to make them more fragile.¹² Structures and routines become ossified around the status quo, reducing exploration of alternatives and responses to disruption. Smoothness and stability reduce diversity; when society appears to be on a steady trajectory, all attention is paid to small variations around the trajectory. When a crisis hits—as it will—the lack of diversity retards flexible responses. As Axelrod and Cohen note, “Even when the sources of external shocks cannot all be identified, a general principle is that the more diverse the elements, the less the chance that they will all be vulnerable to the same kinds of external shock. Diversity in Complex Adaptive Systems not only allows exploration of new options but also provides resiliency against common shocks.”¹³ Moreover, systems aimed at smoothing out disturbances tend to more narrowly constrain the options: they do not simply rule out some possibilities, but channel activity along certain paths. They do not seek to promote optionality as much as induce taking up correct or safe options. As such, diversity of options is reduced, exploration is constrained.¹⁴

⁸ “German Model Leads Way in Job Protection,” *Financial Times*, March 24, 2020.

⁹ As an example, in the 2020 economic collapse the US federal government offered billions in funding to the airlines, 70% of which was in the form of grants if the airlines did not cut staff or routes before October 1, 2020. In the interim period it became clear that the airline industry required a major restriction of its routes and services in light of the drastic reduction in air travel, but while European airlines started before the summer of 2020, US airlines burned cash by supporting services that could not possibly be continued after October 1.

¹⁰ Taleb also endorses limiting the size, concentration, and speed at which disruptions flow through the system. *Antifragile*, chap. 7. Schumpeter, however, points to the benefits of monopolistic practices in his argument that pursuit of short-run efficiency undermines resilience. *Capitalism, Socialism and Democracy*, chap. VIII. Witness the fragility of leveraged bought-out firms in economic downturns.

¹¹ Bosanquet, *The Philosophical Theory of the State*, pp. 189–90. Elster has recently taken over this motto. *Securities against Misrule*, p. 2.

¹² Taleb, *Antifragile*, chap. 6.

¹³ Axelrod and Cohen, *Harnessing Complexity*, p. 108. See also Page, *Diversity and Complexity*, chap. 5.

¹⁴ Recall *The Economist's* recommendation about opening up the Cuban economy (\$18.3).

This leads us to Taleb's notion of the "antifragile": that which is strengthened by stress. "Antifragility is beyond resilience or robustness. The resilient resists shocks and stays the same; the antifragile gets better."¹⁵ In systems with considerable volatility, diversity of responses and strategies abound, and thus continued self-organization is enhanced. The adjacent possible is being thoroughly explored. Such systems—like diverse ecosystems—are far better equipped to respond to the inevitable unexpected shock than systems that have had a long run of smooth sailing. Financial markets are an excellent case. As some complexity theorists have noted, such markets can be highly predictable during certain smooth periods ("pockets of predictability") but are then subject to unpredictable shifts.¹⁶ Attempts to smooth out financial markets by government and central banks can aggravate and radicalize these shifts. Constant volatility encourages diversity: individuals' normal routine is to cope with change in a wide variety of ways and they are always readjusting their risks.¹⁷ Destructive cascades are thus less likely because individuals are accustomed to disruption, searching, and loss-taking, and so are less apt to engage in crowd-following panic behaviors.

23.4. Deontic Rules

Hayek seems to construct strikingly asymmetric criteria for governance and private action.¹⁸ On the one hand, market exploration is based on risk-taking, change, and confidence in self-organizing forces of the market. However, when addressing the problems of governance, Hayek seems to invoke a Burkean fragility thesis. The complex order is a system for social rules whose working we do not understand, and so we should have a reverence for the traditional, evolved rules. But, we might query: If entrepreneurs should be risk-takers and innovators, why are governance changes in the rules so worrisome?

Now interestingly, macro-selection accounts of the market seem to have precisely this conservative upshot. If the interconnected rules and institutions of a society have evolved through strong macro-selection pressures, all we really know is that the system works. We should indeed be cautious about altering its structure, in much the same way that we should be cautious about altering our DNA: given the opacity of, and interconnections among, the rules, there is a high risk that the intervention will be dysfunctional. But the system that has evolved, on Hayek's account, just *is* the market order: the evolved system simply is one in which individuals should be risk-takers within the rules of the market and traditional morality—about which we should not be risk-takers. However, I have rejected the macro-selection account of the Open Society in favor of a strongly self-organized analysis, which includes the self-organization of the rules themselves (§20). Because a rule change takes place within a self-organized system, a change in any particular rule should be the occasion

¹⁵ Taleb, *Antifragile*, Prologue.

¹⁶ Johnson, *Simply Complexity*, chap. 4.

¹⁷ Such cascades are a feature of complex human systems. I have modeled a beneficial cascade in section 20. See Taleb, *Antifragile*, Prologue.

¹⁸ See my "Social Complexity and Evolved Moral Principles."

for further self-organization. Individuals respond to that change much as they would to an endogenous shock, or a change in the behavior of others to which one must reflexively adapt.

The upshot of our account is that, although we would expect a complex order to self-organize in response to a deontic-based law, this is part of the constant adjustments in such a system. Consider, for example, the 1964 Civil Rights Act, prohibiting many forms of racial discrimination. This was primarily a deontic-grounded intervention: such discrimination was morally objectionable on the perspective of most citizens. The aim was not primarily to enhance the system's functioning as in resilience-based interventions (§24.3): by and large it rested on the conviction that such behavior is not to be allowed across a wide range of social interactions.¹⁹ Now the point here is that on a self-organization account, complexity cannot be employed to show that such a change in rules is especially risky or maladaptive. As with any systematic change, we should expect reflexive self-organizing responses that we did not anticipate—such as the fundamental political realignment induced by the Republican Party's Southern strategy appealing to disgruntled Southern whites. Such whites viewed the Civil Rights Act of 1964 as an unwelcome change imposed by the system governor. If desegregation had proceeded more slowly, perhaps in the long term it would have been less disruptive, but that was largely beside the point. The Civil Rights Act of 1964 law had a strong deontic basis in the perspective of most supporters, and its justification was not sensitive to such unexpected reflexive responses—it did not have the goal of promoting system adjustments. Indeed, it induced new conflicts and led to numerous unanticipated and fundamental changes in American political life.

§25 Strategic Dilemmas and Polycentricity

25.1. Global Strategic Dilemmas?

Thus far my concern has been macro governance—system-wide goal pursuit and overall institutional structure. We have considered two dimensions: goals for the entire system, and the institutional framework that structures systematic self-organization. What of our third dimension of governance—macro strategic dilemmas? I doubt there is such a category. All strategic problems are what I call “sectoral”—that is, problems that pertain to a variable or feature (e.g., CO₂ emissions) or a cluster of variables/features (e.g., pollution). We are always concerned with securing cooperation to solve this or that strategic dilemma or challenge, but the existence or direction of the system

¹⁹ To be sure, we must allow for diversity here, too: there is a good case to be made that many businesses in the American South welcomed anti-discrimination laws because they would allow commerce with African-Americans while protecting the business against white punishment for breaking segregation norms. See Muldoon, “Understanding Norms and Changing Them,” pp. 144ff.

is not itself a grand strategic dilemma that needs solving. To be sure, political philosophy has often seen social order itself as a strategic dilemma: governance itself is said to be required to obtain the Pareto-superior outcome of social order rather than war. For Hobbes, without governance “there is no place for industry, . . . no culture of the earth, . . . no knowledge of the face of the earth, no account of time, no arts, no letters, no society.”¹ Perhaps this would be a truly global strategic dilemma. At this juncture in our inquiry, however, the utter implausibility of the Hobbesian program should be manifest. Humans evolved to cooperate: we are a cooperative and networking species. We confront diverse strategic dilemmas, but social order itself is not one of them. Hayek is certainly correct that in this way the Hobbesian program is hopelessly rationalistic, atomistic, and constructivist.

To be sure, both economics and political philosophy have identified “sectoral” strategic dilemmas at the systemic level—defense and pollution abatement are the most obvious examples. However, we need to remember that at this large-scale level, describing, say, defense as a public good confronting us with a collective action problem is a highly stylized “fact.” In truth, many people do not view defense as it is provided as a public good, but think the dangers posed by the military-industrial complex outweigh any security it provides. And recall that to show there is a genuine strategic problem, one must not simply show that each prefers the good to not having it (most people prefer free goods), but that each prefers having it and paying for it to not having it and not paying for it. As elementary as this sounds, it is often forgotten. Many do not prefer paying the cost of curbing CO₂ emissions (+ obtaining the benefits of curbing) to not paying (+ not having). Perhaps their calculations are based on bad evidence, errors in reasoning, or extreme time-discounting, but the fact remains that they do not see themselves in a strategic dilemma. Consequently, a policy to pursue these ends will have a mixed character: to some citizens the policy solves a strategic dilemma and gives them what they prefer, but to others it will be an exogenous aim imposed by the democratic governor. The larger the second group, the more the policy will fall under “sectoral policy goals” (§27) rather than strategic dilemmas. Again, this makes a difference, as reflexivity of responses is a much more serious problem in the former, as many individuals do not concur with the goals of the governor.

25.2. The Pressing Problem-Solving Context and Polycentric Systems

The more we zoom into subgroups, the more likely it is that the relevant population truly shares a strategic dilemma: they overwhelmingly see themselves as facing a strategic problem. One of the lessons we have learned from the work of the Ostroms is that effective joint action is most apt to arise when a group of people face what I shall call a *pressing problem-solving context*, such as when a group confronts a serious degradation of a common pool resource on which it depends. In problem-solving contexts, diversity is reduced because people share (i) a common perception of a problem to be solved, (ii)

¹ Hobbes, *Leviathan*, p. 76 (chap. 13, ¶9). See generally Parsons, *Social Action*, pp. 89–95. For Humean-inspired account along these lines, see Hardin, “The Priority of Social Order.”

an agreement that a range of policies constitutes plausible solutions to the problem, and (iii) a belief that most any of these solutions would be preferable to leaving the problem unresolved. If crime is rising in my neighborhood, my focus is on solving *that* problem; to a significant extent, many of my other diverse aims and goals are bracketed; a crime-fighting community becomes a simpler (less heterogeneous) community, thus reducing the complexity of the public policy problem.

When the problem is, in addition, *pressing*, most individuals in the group believe that solving the problem is sufficiently important that other unintended consequences are not weighty *as far as they are concerned*. Given the inevitable interconnectivity of activities in a complex system, when the group solves their pressing collective action problem there are bound to be other effects, both known and unknown. But when the problem is pressing, the participants will tolerate a significant range of unintended consequences in order to effectively solve the problem. To be sure, insofar as these can be anticipated, they will enter into the problem-solving deliberations, but the critical point is that when an overwhelming majority in group *G* sees a collective action problem as pressing the system becomes, virtually, simpler. Heterogeneity is reduced (because members of *G* share similar goals) and the system's interconnections can be bracketed by the group member (because members of *G* do not care much about them).

The key to the Ostrom-inspired polycentric approach is to, as far as possible, allow problem-solving groups to organize themselves in such a way that the level of governance approaches the optimal public for that strategic problem: just large enough to encompass the stakeholders who perceive a pressing common problem and whose participation is essential to solve the problem. Note that there is a strict limitation on the scope of the aims of governance: to adequately solve a perceived pressing common problem. In the polycentric approach, diverse problem-solving institutions—state as well as non-state—cooperate in forming rule-based efforts to solve shared pressing problems.² Such polycentricity has five major attractions.

- (i) Because the governing unit is focused on a smaller set of problems and common perceptions of solutions, its recursive monitoring task is much easier. We saw earlier that in a reflexive system each participant responds to the judgments of the governor, greatly complicating the problem of governing the system. When the governor is solving shared pressing problems of the group, anticipating the reflexive responses to its decisions is simplified. Note here that when government covers a large and diverse population such that it seeks to solve problems that many participants do not see as problems, this benefit of polycentricity is lost. The governor is then no longer able to anticipate their reflexive response to its decisions, since they do not share its problem-solving orientation.
- (ii) As far as possible, the polycentric program encourages duplication and competition among different polycentric problem-solving institutions. It is often difficult to know what constitutes a good available solution to a problem: the costs and benefits of different institutional schemes is often obscure. Competition among providers in the same public provides a way to decide among different

² See here Vincent Ostrom, "Polycentricity"; Aligica, *Institutional Diversity and Political Economy*.

approaches, and well as for each to learn from each other. Again, because the competitors are seeking to solve the same or similar problems in the same broad norm context, institutional learning can occur (that is, the “institutional transplant” problem is avoided).³

- (iii) Polycentric institutions are in an important sense themselves part of social self-organization. As we see in Elinor Ostrom’s work, they often arise within a self-organized network of relations; they certainly are not top-down governance imposed on the self-organized system.⁴ They form out of the self-organized networks and common perceptions of inadequacies in them and the need for governance. Polycentric systems thus provide space for norm exploration: groups experiencing perceived unsolved (or badly solved) collective action problems seek to resolve them within the context of their current social networks. This may require altering the norms of the interactions.⁵
- (iv) The social conflict that heterogeneity can engender is transformed in problem-solving contexts into a more cooperative inquiry looking for better solutions. Once politics is conceived in terms of inquiry into the best solutions to a common problem, we can draw on results such as those of Lu Hong and Scott Page, which show how diverse groups possess enhanced problem-solving capabilities (for some explanation, see §29.2).⁶ It is important that Page’s diversity theorems are about problem-solving contexts: when we have identified a common problem and have agreed on what would be a good solution, then Hong-Page dynamics can get going. Because democratic polycentric citizenship is about collective problem-solving, the stage is set for diversity to assist in social searches for better solutions.⁷ Pressing collective action problems thus have something of the perfect mix of homogeneity and heterogeneity for Hong-Page dynamics: an agreement on the problem, its importance, and a general concurrence on what would constitute good solutions with heterogeneity of perspectives, toolkits, and cognitive resources, so that the solution space can be more adequately explored.
- (v) Because polycentric self-organization does not commence with a certain pre-defined group (say, a national state), it can adjust its boundaries to encompass all those who share the pressing problem orientation. Rather than commencing with a preferred unit (which is apt to be highly diverse and complex) as the focus for all policy, different publics form at different levels in response to various collective action problems: the boundaries of the public seek to track the simplification of the social problem induced by the pressing problem orientation, which also enhances the ability to actually solve the problem. It is of the first importance that polycentrism is not understood in terms of the autonomy of local or small

³ See Vincent Ostrom, Tiebout, and Warren, “The Organization of Government in Metropolitan Areas”; Alga, Boettke, and Tarko, *Public Governance and the Classical Liberal Perspective*, chap. 7.

⁴ See Elinor Ostrom, *Governing the Commons*; Lane, *The Complexity of Self-Government*.

⁵ See Colander and Kupers, *Complexity and the Art of Public Policy*, pp. 28, 181ff.

⁶ See, e.g., Hong and Page, “Problem Solving by Heterogeneous Agents”; Hong and Page, “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers”; Page, *The Diversity Bonus*. See also Müller, *Political Pluralism, Disagreement and Justice*, chap. 7.

⁷ See Landemore’s example of collective decision-making in New Haven about preventing muggings, which provides an example of the Hong-Page theorem at work in a pressing problem-solving context. *Democratic Reason*, pp. 100–2.

communities. For any given problem, the proper size of the democratic self-governing public ranges from the neighborhood to the globe. At the same time, as we scale up the public, we can expect that the percentage of the population who perceive it as a common pressing problem will decrease. Governance based on strategic dilemmas thus appears most effective at the meso to micro levels.

Polycentricity's partial reconciliation of self-organization and self-governance leads, on Paul Dragos Aligica's analysis, to a revised conception of democratic citizenship. Especially prominent in his account is the task of the "public entrepreneur," who takes the lead in identifying, and showing the importance of, potential collective problem-solving contexts. That a group confronts a collective action problem does not mean that the problem is obvious or is already recognized as pressing. The public entrepreneur takes a leading role in mobilizing recognition of the problem and ways to solve it, which includes providing the contexts for discussion and exchange of information.⁸

25.3. Polycentric Self-Governance

In section 22.4, I followed the traditional game theoretic modeling of public goods, club goods, and common pool resources problems. Like many models, when used properly it is enlightening, but it can be distorting unless we appreciate how much it misses. Solving these problems is not simply a matter of "equilibrium selection," as if all that is required is for a governor to declare, "Cooperative strategy A will be followed!" (and perhaps punish defectors). One of the great contributions of the Ostroms was to focus on the way in which solving strategic problems is a matter of institution creation: for a group to solve its common pool resource problem, for example, it must formulate a self-governance structure that requires both first-order rules about how to effectively and equitably manage the resource and second-order rules about how to reflexively evaluate the functioning of those first-order rules.⁹ First-order rules include those identifying the boundary of the public involved, prohibitions that trim the set of acceptable options, and permission rules that specify when governors are allowed to act.¹⁰ Solving strategic dilemmas thus requires establishing the rules of the public good, club good, common resource, or coordination games.

Because solving real-world collective action problems is a matter of institutional development, they could not be solved if we were simply the sort of self-interested maximizers of economic theory. All rule system are porous and present myriad opportunities for cheating. As the Ostroms stressed, self-governance takes place among rule followers (§11.1).¹¹ It is our ability to be guided by norms that allows us humans

⁸ See *Public Entrepreneurship, Citizenship and Self-Governance*, chap. 2, where Aligica demonstrates that the entrepreneur can build on different preferences for public goods, helping to show how diversity of preferences can cause convergence on outcomes, not simply divergence. See also Aligica, Boettke, and Tarko, *Public Governance and the Classical-Liberal Perspective*, esp. chap. 5.

⁹ Elinor and Vincent Ostrom, "The Quest for Meaning in Public Choice," pp. 8ff; Elinor Ostrom, *Governing the Commons*, pp. 50ff.

¹⁰ Elinor Ostrom, "An Agenda for the Study of Institutions."

¹¹ *Ibid.*, p. 100.

to self-organize to avoid strategic dilemmas which, to the rational choice theorist, seem insoluble without the coercive state. The Bloomington School's work repeatedly demonstrated people's ability to self-organize self-governance systems to solve collective action problems.¹² The sort of self-organization of individual norm networks that I formally modeled in section 20 has been shown through extensive fieldwork to be a critical way in which groups actually organize their own governance systems.¹³ "Polycentric systems are open systems that manifest enough spontaneity to be self-organizing and self-governing."¹⁴ The great advantage of self-organizing self-governance is that the governor does not seek to impose top-down directions on the self-organized norm network: governance itself flows out of these norm networks. Whereas top-down governance must strive to cohere with the self-organized norm networks (§22.2), self-organizing self-governance grows from these norms to institute rules of governance to address pressing shared problems.

Polycentric problem-solving publics are no panacea. I have stressed that when a group sees a pressing problem, they tolerate a great deal of unforeseen consequences in pursuit of a solution to their pressing problem, thus practically reducing the relevance of interconnections. As far as they are concerned, it is pretty simple to evaluate whether their self-governance secures its goals: it adequately solves and manages the pressing collective action problem—a problem that often was created by their own choices. But the tight coupling that characterizes complex systems still exists, and those who do not share the pressing problem are apt to be impacted by their neighbors' solutions to their collective action problems. A clear example is what Tenner has described as "rearranging effects" where one group's solving its problems pushes the problem to another group. Thus, for example, in response to rising ocean levels and beach degradation, some communities have constructed breakwaters, which have had the effect of disrupting complex sand flows in the ocean, resulting in starving the sand required for beaches further down the coastline.¹⁵ Sometimes these problems can be solved by expanding the problem-centered public to include the wider group, but because these rearranging effects are often not perceived, or the costs of solving the wider problem may be high, groups solving their own problems may continually create problems for others.

The critical point is that, in and of itself, this is not an objection to polycentric governance. It is the very nature of self-organizing complex systems that one group's activities create externalities—both positive and negative—for others. When we are tightly linked in a complex system, each person's action always is changing the costs and opportunities for others. The ideal of actions that internalize all costs and benefits is an artifact of an economic model of pure static efficiency. Thinking about Robinson Crusoe and Friday may be a helpful idealization in some contexts, but not when thinking about the complex Open Society. The systemic rules of the game in a complex order seek to thwart the most harmful or inequitable sorts of externalities

¹² See Aligica and Boettke, *Challenging Institutional Analysis and Development: The Bloomington School*.

¹³ See Elinor Ostrom, *Governing the Commons*.

¹⁴ Vincent Ostrom, "Polycentricity: The Structural Basis of Self-Governing Systems," p. 58. For a converging analysis from a different starting point, see Lane, *The Complexity of Self-Government*.

¹⁵ Tenner, *Why Things Bite Back*, p. 93.

from flowing over to others (§18.4), but attempts to wall off one group's actions from another's options would undermine the flexible self-organization on which the Open Society depends. To be sure, when especially serious and known, these externalities may well require intervention by a higher-level governor, but the aim is by no means to eliminate negative externalities.

§26 Meso-Level Goal Pursuit

26.1 Resurrecting Goal Pursuit in Subsystems?

I have agreed with Hayek that macro guidance of an entire complex order is a deeply implausible governance project. What, though, if we zoom down to smaller subsystems? Certainly this simplifies some aspects of the problem: Arizona is less complex than the United States, Wales less complex than the United Kingdom, and Denmark less complex than the European Union. Hence it may be hoped that some sort of subsystem macro planning and guidance are still possible. However, this too is unlikely. For one, whatever gains in simplicity are attained by zooming down to the meso level are almost surely offset by the fact that more of the influences on the (sub)system are exogenous. Leaving aside the hope of high levels of decomposability and modularity, while Wales may be simpler than the United Kingdom, a great deal of what happens to Wales depends on what is happening in the rest of the United Kingdom and the world. We might say that the endogenous variables have a simpler relation to each other, but much more of the action depends on exogenous variables, defeating overall subsystem plans. France's long-standing efforts to plan its economy were somewhat more feasible before the euro than they are now.¹

And in any event, it is easy to underestimate the complexity of large subsystems, such as great cities. That planning to guide a meso-level system is in constant conflict with self-organized diversity is the theme of Jane Jacobs's classic *The Death and Life of Great American Cities*. Jacobs, back in 1961, insisted:

Cities happen to be problems in organized complexity, like the life sciences. They present "situations in which a half-dozen or even several dozen quantities are all varying simultaneously and in subtly interconnected ways." Cities, again like the life sciences, do not exhibit one problem in organized complexity, which if understood explains all. They can be analyzed into many such problems or segments which, as in the case of the life sciences, are also related with one another. The variables are many, but they are not helter-skelter; they are "interrelated into an organic whole."²

¹ I return to some questions of meso policy and the European Union in §28.2.

² Jacobs (quoting Warren Weaver), *The Death and Life of Great American Cities*, p. 433.

As she depicts in detail, cities depend on the complex life and diversity of streets, neighborhoods, mixes of residential and commercial, families and singles, norms of anonymity, observation, and accountability. Their dynamics can be understood, but cities cannot be guided or planned in any overall way. Tremendous efforts were made in the 1950s–1970s to do so—most famously the garden city movement, which sought to radically decompose this deeply diverse complex system, dividing up residential from commercial, children from adults (playgrounds for children, streets for adults), only truly mixing children and adults at home and in parks.³

The two major variables in the Garden City concept of planning were the quantity of housing (or population) and the number of jobs. These two were conceived of as simply and directly related to each other, in the form of relatively closed systems. In turn, the housing had its subsidiary variables, related to it in equally direct, simple, mutually independent form: playgrounds, open space, schools, community center, standardized supplies and services. The town as a whole was conceived of, again, as one of the two variables in a direct, simple, town-greenbelt relationship. As a system of order, that is about all there was to it.⁴

Garden City-inspired planners moved residences back away from street, surrounded by green areas that often became highly dangerous crime zones, while removing residential life from the street life also increased street violence and predation. The parks—intended as a refuge for children—themselves often became unsafe and drug infested. As Jacobs shows again and again, these were not largely exogenous shocks, but the result of seeking to shape the life of a complex system to conform to the ideal of city planners.⁵

26.2. Homogeneity within Diversity: A Liberal Plan for Closed Societies

Utopians and communitarians are thus faced with a dilemma: we live in an increasingly diverse society, which induces a degree of complexity that frustrates the pursuit of collective goals or, as philosophers would say, shared views of the good life. One perennial and still popular response is to seek to stamp out diversity. A more sophisticated and “liberal” response, which is often confused with polycentricity, is an order based on politics “with more like-minded people.”⁶ On this view of governance under

³ Having lived in a garden city designed from scratch—Canberra—I can testify to the empty wasteland that the “city center” became by 7:00 p.m. This clearly was not the result of exogenous shocks but was, alas, the aim. It took years for the city to mitigate the garden city plan, seeking to undo the decomposability thrust on it by planners and inducing a modicum of self-organization.

⁴ Jacobs, *The Death and Life of Great American Cities*, p. 435.

⁵ The “right” as well as the “left” often seeks out simplicity and uniformity. Perhaps one of the most worrying features of libertarianism is its fascination with the uniformity of the gated suburban community. It was the self-organized diversity of the great European cities that gave rise to liberalism. See further my “A Libertarian Alternative to Liberal Justice.”

⁶ See Müller, *Political Pluralism, Disagreement and Justice*, chap. 9.

diversity, like-minded individuals band together in meso-level and local communities within a broad framework of liberal rules.⁷ This would allow for subsystem planning, since within each subsystem diversity is greatly reduced: it is a diverse federation of homogenous communities. Nozick famously presents a “framework for utopia”—in which, by a combination of planning and evolutionary processes, different communities build their own versions of an ideal society. All these communities exist within a broadly liberal framework ensuring basic rights.⁸ Chandran Kukathas proffers a rather different vision of a “liberal archipelago.” Here again, different groups pursue diverse ways of living, but within a much more modest framework, based essentially on a person’s claim to exit communities.⁹ The similarity to Ostrom-inspired polycentric government—especially the stress on diverse institutions and meso-level governance—is manifest. Yet the differences are also stark. As I have presented it—in the spirit of Elinor Ostrom’s work—polycentricity is a method of problem-centered governance: its grounding in a pressing problem context inherently limits the focus of any polycentric governance unit to solving a set of strategic dilemmas among its members. Because of this, Ostrom-style polycentric governance induces a crisscrossing of governance units at multiple levels, depending on the problems confronting some public. As Elinor Ostrom reiterated, the scope of the governance unit depends on the scope of the problem. In contrast, the proposal under consideration focuses on some single governance unit—a “polity,” “community,” or even “culture”—which consists of a group bound by reinforcing norms on many matters. These “thick” communities are related only by a “thin” overall framework of interactions (a minimal state for Nozick, far thinner for Kukathas). Whatever the attractions of such proposals to some, we have good reasons to be skeptical that they are frameworks for open societies. As we saw in section 10.3, when diverse populations split into isolated groups, we should expect either animosity or, at best, a sort of practical relativism. This, in turn, will tend to lessen the interest in reconciliation with the moral views of others, and so decrease the likelihood that there will be convergence on the justification of an impartial framework. Because of this, the overarching framework in which the like-minded communities relate is apt to be seen as unjustified by most. After all, the other group will often have a very different set of rules emanating from a very different set of ideals, and there may well be little in the socially eligible set between them. As the Ostros observed, “a highly-fragmented political system *without* substantial overlap” is especially prone to conflict.¹⁰

For Nozick, an attraction of such communities is that they can pursue their ideals and build a type of society. Whereas a diverse society will share few if any ideals or ordering of values, these societies can collectively pursue their ideals. Perhaps, but only

⁷ For a comprehensive history and analysis of this strand in liberalism, see Levy, *Rationalism, Pluralism and Freedom*.

⁸ Nozick, *Anarchy, State and Utopia*, chap. 10.

⁹ Kukathas, *The Liberal Archipelago*.

¹⁰ Vincent and Elinor Ostrom, “Public Goods and Public Choices,” p. 96. Of course, nothing is inevitable: it seems that such systems can be (at least for a time) managed by agreement among elites. See Lijphart, *Democracy in Plural Societies*. Cf. Müller, who thinks that tensions would be reduced by separation. *Political Pluralism, Disagreement and Justice*, chap. 9.

if they are strikingly uniform. If we suppose that, absent strong use of state and social coercion, people will develop different perspectives on their shared ideals, we can expect most of the same problems—and more—that thwarted the pursuit of systemic goals to thwart these meso-level communities' quests.¹¹

Consider, then, a reasonably appealing version of this brand of liberal pluralism. The communities tend to be devoted to different ideals, but they are not oppressive: people come to disagree about the specific nature of the ideals and the best way to go about securing them. And while the communities are distinct, they are not isolated in the archipelago: they communicate and trade. Given this, at the community level the governor is again seeking to control behaviors that are greatly impacted by events outside its subsystem. If s_i is a system of S , the elements of s_i are connected to elements in S outside of s_i , which will greatly impact the behavior of s_i itself. If the governor cannot control these elements, the job of controlling s_i is hopeless unless the system is *highly* decomposable (§17.3). If a governor is pursuing an egalitarian social policy of equity over efficiency, but is linked in international trade with foreign firms stressing efficiency, the egalitarian policies will be impacted. Because the course of autocatalytic diversity is to weave together subsystems, community governors seeking control must self-isolate (they will not build bridges uniting the islands in their archipelago). Within the community, people will come to disagree about the best way to pursue their ideals, so responses to the directives of the governor will again be varied, and reflexive: people will react to what they see as the pattern of ideal pursuit which their joint decisions have yielded. Unless the communities enforce isolation and homogeneity we cannot, after all, expect success in the collective meso-level pursuit of their ideals. All in all, the proposal looks much more like a flight to closed societies than an alternative model of the Open Society.

§27 Sectoral Policy

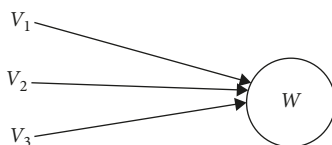
27.1. Many-to-One Influence

Although a good deal of political philosophy advocates guiding the entire system (or subsystem) along some planned path, most policy-based governance is more modest. Policy typically seeks to maximize the values of a type of social phenomenon—what I call a “sector”—such as economic growth, health, or education. Each of these is actually a set of closely linked variables (growth may concern GDP growth, employment growth, manufacturing growth; health involves longevity, days without chronic disability, rate of serious disease, etc.), each of which can be measured in a wide variety of ways. Because each of these phenomena are multidimensional, problems of trade-offs

¹¹ I have sketched the way Owenite communities fractured on different interpretations of their ideals of liberty and equality. See my *The Tyranny of the Ideal*, pp. 90–3.

and complexity can arise among the different elements of a “policy.” Does increasing employment growth require a decrease in manufacturing growth, and how will that impact GDP? To initially make our analysis tractable, I will set aside for now these important issues and focus on sectoral goals as if, say, “economic growth” and “education” are essentially each single goals. This is a strong idealization that we will later relax, but it allows us to initially focus on some fundamental matters.

The core tools of policy analysis are multiple regression and vector analysis, based on what we might call a “multivariate” analysis. To greatly simplify, these methods recognize that the target policy goal (W) is affected by multiple variables $\{V_1 \dots V_n\}$. Again greatly simplifying, the aim, say, of multivariate analysis is to apportion causal influence of the various elements of $\{V_1 \dots V_n\}$ —to determine the degree of variance in the value of W explained by, say, variance in V_n . This, then, is a many-to-one analysis: taking many factors, we seek to explain their impact on our target variable. Now clearly this method only yields enlightening results when the set $\{V_1 \dots V_n\}$ is not too



Display III.6. The multivariate view of the world.

large, and accounts for much of the variance in W . In a decomposable system we might discover that $\{V_1 \dots V_n\}$ accounts for a high proportion of the variance in W : we can develop an accurate hypothesis of the causal influence of each member of $\{V_1 \dots V_n\}$ on the variance in W . The basic multivariate model is depicted in Display III.6.

Policy scientists uncover significant multivariate relations that indicate the causal factors influencing a target variable (W) at some time. At first glance, this may seem inconsistent with complexity, but even highly complex systems can be predicted for certain intervals. The best example is weather, which is universally acknowledged to be a highly complex system (think again of the Butterfly Effect, §23.2). The weather forecasts for tomorrow are accurate, three days out still reasonably good, next week far less so, and at two weeks we are essentially predicting long-term averages. Models of weather forecasting are thus characterized by *error inflation*: the ability to predict W on the basis of $\{V_1 \dots V_n\}$ radically decreases as time proceeds. The familiar “cones” in projected hurricane paths widen as time proceeds, thus modeling error inflation. In such cases, we can have analysis at time t_1 that effectively explains most of the influences on our target W , be it the weather tomorrow or growth rates this quarter. But W occurs within a complex system; while at time t_1 $\{V_1 \dots V_n\}$ may well be the most influential factors, as time proceeds, other elements of the interconnected system can swamp the influence of $\{V_1 \dots V_n\}$.

Not only can the causal influence of $\{V_1 \dots V_n\}$ be dissipated in a complex system over the medium term, but as the system reorganizes, what was previously a clear causal connection may itself change. In complex systems what was once a causal

regularity may disappear. We might call this the *Disappearing Causal Laws Effect*.¹ For a time there was, for example, a close relation between employment and inflation (at full employment, inflation should increase markedly) or between employment levels and wage growth (at full employment, wages should increase at a reasonably rapid rate); recent experience indicates that these relations are, at best, now muted. This should be expected: in a complex system, the upshot of any specific set of variables is greatly impacted by developments in other parts of the system; since the system is nonlinear and path dependent, apparently clear causal relations can appear, then disappear.²

As I have been stressing, a complex system at the meso level can be part of a macro-level complex system. Weather, for example, is a complex system embedded in a macro-level complex climate system. When analyzing the macro climate system over the short-to-medium term, weather systems can be treated as discrete data independent points, and they can be averaged to predict, say, temperature or rainfall over the medium term. As noted earlier, after ten days to two weeks, weather prediction is essentially medium-term averages. But the climate system itself is complex over much longer time scales, and as climate modeling indicates, it is nonlinear. Given that, like all complex systems, climate is more predictable over the short term than the long term.

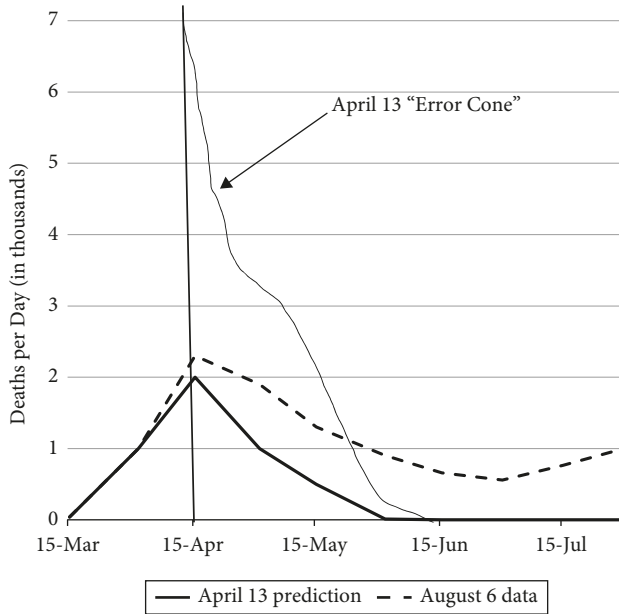
I have been focusing on models that seek to understand the causal relations in many-to-one contexts by understanding the relations between the variables, as in multiple regression analysis. Alternative models may abstract from many causal factors, and simply seek to fit the curve of past data and extrapolate on the basis of certain assumptions. The Institute for Health Metrics and Evaluation (IHME) model of COVID deaths to a large extent utilized this method, employing observed death rates from COVID-19, incorporating information about social distancing, other protective measures, and including uncertainty intervals.³ Display III.7 contrasts the model's predictions on April 13, 2020, and the observed deaths in the United States as of June 15, 2020.

Display III.7 simplifies the data, smoothing out daily and weekly variations, and taking approximately fortnightly data points. We see that in the roughly three-month time period not only did the predictions change a great deal, but by mid-June the estimated daily deaths were outside the rather expansive uncertainty interval of the April 13 prediction. Indeed, we observe the classic nonlinearity of complexity, as (perhaps because of the reflexivity of citizens), the actual curve changes direction in July. None of this shows that we cannot predict many-to-one relations, but again we see how the predictions become increasingly unreliable as error inflation sets in—in this case, within three months they were outside the predicted error zone. By July 1, 2020, deaths were beginning to trend upward.

¹ Gillian Trent, "Weird Things Are Happening to the Market," *Financial Times*, April 5, 2019.

² This is one of the reasons why big data is no panacea. If the parameters change, regularities that were once law-like can suddenly cease to hold. Cf. Einav and Levin, "Economics in the Age of Big Data." This also could be an explanation for the instability of economic findings. See Goldfarb, "Now You See it, Now You Don't: Emerging Contrary Results in Economics."

³ A weakness of such models is that their predictions can be radically off when an unexpected tipping point is encountered that switches the system to a very different trajectory. See §27.4. The IHME model evolved, integrating other approaches. As the model developed, it no doubt became more accurate.



Display III.7. IHME modeling of United States Covid-19 deaths, April 13 and August 6, 2020.

Note: Data accessed April 13 and August 18, 2020, from <https://covid19.healthdata.org/united-states-of-america>. Numbers are rounded.

It is critical to appreciate that there is no *a priori* reason why long-term prediction should be less reliable than short-run prediction. Quite the contrary. Let us say that a *Complicated Determinate System* is one that is characterized by stable independent, causal nexuses among discrete sets of variables with a great deal of stochastic (random) influences. In such systems, we would search out many-to-one causal relations such as in Display III.6, but our task would be complicated by a great deal of noise and random events that are constantly impinging on the long-term, stable causal relations. This is the sort of system in which Durkheim's model of statistical-driven social science is most appropriate.⁴ Here long-term prediction should be much more reliable than short-term, since over the long run the noise will cancel out, leaving behavior to be determined by the stable causal relations. A statistician's prediction of the Minnesota Twins' play tonight will be far less accurate than her prediction of their play over the next three months. Complicated Determinate Systems have long-term predictability and short-run volatility: over the long run, the true causal influences tend to swamp short-term flukes. Marx thus thought he could predict the broad sweep of history, not this year's events: the fundamental causal factors in the end will

⁴ See Durkheim, *Suicide*, esp. Book I, chap. 3; Book II, chap. 1.

win out. Complex systems, in contrast, tend to have short-run predictability and long-run unpredictability: the coupling of elements of the system ultimately swamps a set of short-run dominant factors.

27.2. A Tale of Two Tetlocks?

If social systems are primarily Complicated Determinate Systems, social prediction should typically display higher long-term predictability than near-term predictability; if they are generally complex, short-run predictability should tend to be much better than long-term prediction. Now in deciding this question we must avoid trading favorite stories of predictive successes and failures. In speaking for the last twenty years to philosophical audiences about the difficulty in predicting complex systems, I am invariably confronted by disbelief, as members of the audience recall favored cases that were predicted accurately (and for which intervention succeeded).⁵ If the same philosopher confronted a gambler who said he could recall cases where his roulette system beat the house, she would scoff at such blatant confirmation bias, but when it comes to our ability to predict social events, confirmation bias rules.⁶ Thus the great importance of Philip Tetlock's project to formally measure and test predictions of social phenomena.

Tetlock's work on prediction can be broken into two phases. In the first phase, summed up in *Expert Political Judgment*, Tetlock studied the ability of political and economic experts to predict, among other things, economic performance (growth rates in GDP, inflation, unemployment rates) as well as political developments.⁷ The focus was on predictions five and ten years out. Tetlock asked experts in history, political science, and economics to predict future events and the movement of key variables. Would key variables go up (in both the medium term and the long term), go down, or stay the same? The good news was that in their field of expertise, political and economic experts did better than undergraduates at predicting future events. Unfortunately, that was really about all the good news in this study. Experts did not do significantly better than what Tetlock calls "dilettantes"—people who regularly read *The Economist* or *The New York Times*. Tetlock distinguished two criteria of a good prediction: discrimination (how precise the prediction is) and calibration (how accurate the prediction is). Someone who always makes very general predictions (for example, "a 33% chance of a downward movement in an index") would make a number of accurate predictions (by chance they would be correct one-third of the time), but they would score low on discrimination; someone who predicts "an 80% chance the variable will fall" aims at a precise prediction, but of course they may sacrifice accuracy: they are more apt to go wrong since they are trying to give a precise prediction.

⁵ These are almost always linked to ideology. The libertarian recalls when opening trade promoted growth, the progressive recalls the New Deal.

⁶ Taleb hypothesizes that we store past information in causal narratives (*The Black Swan*, chap. 4); as we recall it, the world strikes us as much more determined than it is. Such hindsight bias is well documented. See Tetlock and Gardner, *Superforecasting*, p. 183.

⁷ Tetlock, *Expert Political Judgment*, esp. chap. 2

On the discrimination measure—how precise the predictions are—the experts and dilettantes would beat a random procedure deciding between “variable will go up,” “variable will go down,” or “variable will stay the same.” Unfortunately, the random procedure beats the dilettantes and experts on the accuracy score.

All this is about comparative performance. What about absolute performance? Experts were better, I have said, on the discrimination dimension—they made more precise, if less accurate, predictions than a random process. How good were they? The better half of the expert group predicted a meager 18% of the variance, the less good group about 14%. An average of about 16% of the variance was accounted for by expert prediction.⁸ Tetlock called the better group “foxes,” who were cognitively open and fallibilistic, and the lower group “hedgehogs,” who were more apt to dismiss evidence discordant with their theories. Based on these findings, Tetlock was forced to concede that crux of the skeptical hypothesis (which he relates to complexity theory): expert prediction and guesswork were essentially the same. It is important that Tetlock is not himself a skeptic. He is hopeful that we can improve public policymaking, so he focused on the difference within the expert group, looking at which sorts of experts tended to do better.

Tetlock seemed stung that his work was interpreted as highly skeptical of expert prediction (though in my view it is hard not to draw that conclusion).⁹ In any event, in the second phase Tetlock developed a more sophisticated way to measure predictive success,¹⁰ and collected a very large data set of over one million judgments about the future. These judgments focused on questions such as:

- “Will Russia officially annex additional Ukrainian territory in the next three months?”
- “In the next year, will any country withdraw from the eurozone?”
- “Will North Korea detonate a nuclear device before the end of this year?”
- “How many additional countries will report cases of the Ebola virus in the next eight months?”
- “Will India or Brazil become a permanent member of the UN Security Council in the next two years?”
- “Will NATO invite new countries to join the Membership Action Plan (MAP) in the next nine months?”
- “Will the Kurdistan Regional Government hold a referendum on national independence this year?”
- “If a non-Chinese telecommunications firm wins a contract to provide Internet services in the Shanghai Free Trade Zone in the next two years, will Chinese citizens have access to Facebook and/or Twitter?”¹¹

⁸ Ibid., pp. 76ff.

⁹ Tetlock and Gardner, *Superforecasting*, p. 4.

¹⁰ Tetlock employed “Brier” scores, which measure “the gap between forecasts and reality, where 2.0 is the result if your forecasts are the perfect opposite of reality, 0.5 is what you would get by random guessing, and 0 is the center of the bull’s eye.” Some of his superforecasters ended up with Brier scores of .14. Ibid., p. 93.

¹¹ Ibid., p. 2.

These are examples of Display III.6's multivariate causation model; a discrete event impacted by a variety of factors. In this phase Tetlock conducted a tournament and uncovered "superforecasters"—people who scored high in predictive ability, and whose performance could be improved.¹² These superforecasters had much in common with his earlier "foxes" (they were in some ways superfoxes).¹³ At first glance, it might seem that the results of this second phase of research overthrew the unwelcomed skeptical implications of the first. Crucially, however, while the first phase focused on five- and ten-year-out predictions, the "superforecaster" phase considered much shorter-term predictions, "with the vast majority of forecasts extending out more than one month and less than one year."¹⁴ Accuracy still drops to around chance from three to five years out. Tetlock's two phases are thus consistent: if we consider social predictions five or ten years out, predictive power approaches nil; in the shorter term, gifted predictors, who are high in cognitive openness and fallibility, can do quite well. As Tetlock himself notes, this is what we would expect from nonlinear, complex systems.¹⁵ This supports the view that the sort of non-actuarial social developments that policymakers are most interested in are better analyzed as parts of a complex system than a Complicated Determinate System.

Tetlock's study also provides evidence that teams of diverse predictors can outperform individual predictors.¹⁶ Tetlock's conclusions converge with Page's results on the importance of diversity in solving problems. Especially important, but often overlooked, is the Diversity Prediction Theorem, according to which, in a diverse group, aggregating predictions from different models outperforms the average of the individual models. In short, *collective error = average individual error – predictive diversity*.¹⁷ Page has shown that even if model A predicts twice as well as model B, aggregating their predictions can improve upon model A's predictions alone (see further §30.1).¹⁸ We thus encounter a predictive version of our important finding from section 20: diversity can be the best response to coping with complexity. The importance of this is difficult to overemphasize. When confronted with the problems of diversity and complexity, an initial response is to simply homogenize and seek guidance from the best expert. But counterintuitively, the best response is very often to encourage more diversity and disagreement, which can give us better solutions. A great value of more formal models of diversity is that they challenge our strong intuitive biases toward homogeneity as the basis of both order and expertise. I return to this important lesson in section 30.

¹² Taleb is skeptical of "superforecasting"—for Tetlock's reply to Taleb, see Tetlock and Gardner, *Superforecasting*, pp. 237ff. As I argue in the text, even if we follow Tetlock, the evidence is still very strong that we are dealing with complex, nonlinear phenomena.

¹³ Tetlock employs Kahneman's dual-system theory, distinguishing System 1 (fast) and System 2 (slow) cognitive processing, arguing that the superforecasters do a superlative job at monitoring their System 1 judgments. *Ibid.*, pp. 33ff, 231ff. See also Kahneman, *Thinking, Fast and Slow*.

¹⁴ Tetlock and Gardner, *Superforecasting*, p. 17.

¹⁵ *Ibid.*, pp. 12ff.

¹⁶ *Ibid.*, esp. chap. 9.

¹⁷ For explications, see Page, *The Difference*, pp. 205–12; Wagner, Schneider, and Chen, "The Wisdom of Reluctant Crowds." For an extended example, see my *Tyranny of the Ideal*, pp. 261–5.

¹⁸ Page, "Not Half Bad: A Modest Criterion for Inclusion."

27.3. Manipulation and Revenge

Tetlock's findings give reason to conclude that *exceptional* forecasters can predict social indicators and events in the one month-to-three-year range, and in some cases up to five years. (I stress "exceptional" here because Tetlock's studies also tend to indicate that the most famous and highly credentialed policy analysts are not good predictors). We need not fully endorse his enthusiasm for "superforecasting," but there is good reason to suppose that reliable near-term prediction of discrete variables and events in complex social systems can be achieved. Again, this is entirely consistent with complexity, nonlinearity, and error inflation. And it supplies the predictive basis for short-to-medium-term highly focused sectoral policy. The first thing a governor wants to know in devising a policy is where the target phenomenon is already headed. But it is only the first thing. In addition, policymakers must possess a "manipulative" capacity (§23.1). To slightly revise a common quip, "everyone predicts the weather, but no one does anything about it." And that is because even good short-term predictive power is distinct from possessing the control levers needed to alter the variables in the desired direction.

These control levers can be most difficult to find. In his review of technological innovations as well as medical and environmental policies, Tenner introduces the idea of a *revenge effect*. Revenge effects are not simply unintended consequences, in which our manipulation to improve the value of a target variable has unintended effects on other variables. Rather, a revenge effect occurs when our manipulation affects the target itself in negative ways.¹⁹ Tenner provides hundreds of pages of examples of such revenge effects. For our purposes, three sorts of revenge dynamics are especially important.

The Paucity of the Policy Toolkit: Only Having Hammers When You Need Scalpels

In many cases, a governor can have good knowledge of the causal variables, but lacks an instrument that allows it to target those causal variables and only those. Consider, for example, the Exxon *Valdez* oil spill off Alaska in 1989. In its massive and expensive cleanup operation, Exxon's goal was to remove the oil from the beaches in a cold climate. Its chosen method was high-velocity hot water pumps to liquify and drive away the oil. The result, as eventually evaluated by the Hazardous Materials Response Unit of the National Oceanic and Atmospheric Administration (NOAA), was widespread deleterious consequences for the beaches: "The treatment scalded the beach, killing many organisms that had survived the oil, including some that were little affected by it. It also blasted off barnacles and limpets. And it drove a mixture of sediment and oil down the beach face, depositing them in a subtidal area richer in many forms of marine life—and one where there hadn't been much oil."²⁰ The point is not that Exxon

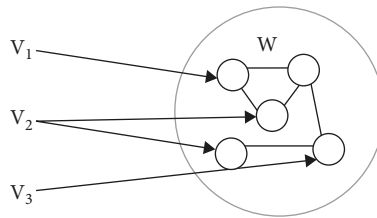
¹⁹ Tenner, *Why Things Bite Back*, p. 7.

²⁰ NOAA Report quoted in *ibid.*, p. 90.

was behaving badly to save money—this was an expensive operation, and doing less probably would have been more effective—but it was widely seen to be an imperative to fix the problem yet the tools available were far too blunt. It is also important that policymakers often are unaware of whether they possess scalpels or hammers. During the 1962 Cuban Missile Crisis there was a faction that pressed Kennedy to order in “surgical strikes” against the Russian missiles, but it was a constant question among the decision-making group just how “surgical” the strikes would be (i.e., how many Russian advisors would be killed).²¹

The Target as a System: Helping Here and Hurting There

Let us relax our assumption that the target variable is singular (\$27.1) and allow that the target phenomenon itself consists of a system of variables. Instead of Display III.6, we have something closer to Display III.8.



Display III.8. A target system.

In such cases, the governor’s manipulation often positively impacts one part of the target system to the detriment of another. Examples of revenge effects in such systems abound, as a successful intervention on one part of the target led to unwelcomed changes in others. In the early 1970s, utility companies in the Midwest of the United States constructed giant smokestacks (some 1,000 feet high) to meet new local clean air standards; this was successful in decreasing particulate pollution around the plants, but had the effect of pushing pollution into the upper atmosphere, contributing to acid rain in the Northeast.²² Efforts to rescue sea otters from oil spills have had the effect of transferring sick and stressed animals to new populations, in which they spread diseases.²³ Tragically, there is evidence that the aim of freeing up hospital beds to cope with the expected influx of COVID-19 patients led hospitals to discharge patients into care homes, who subsequently infected clusters of elderly at-risk residents.²⁴ The aim of improving medical care and coverage has encouraged people to undergo more procedures, but all hospital stays have a non-negligible risk of harm

²¹ Kennedy, *Thirteen Days*.

²² Tenner, *Why Things Bite Back*, p. 86.

²³ Ibid., p. 90.

²⁴ *Financial Times*, May 15, 2020. Many of these beds were ultimately not needed at that time.

through error. One study concludes that one patient in a hundred is negligently injured; as people enter hospitals for increasingly minor procedures, the number of those “unnecessarily” injured increases. According to another estimate, avoidable injuries in hospitals cause twice as many deaths each year as highway accidents.²⁵ Closer to home (for many of us), it has been conjectured that universities’ reasonable policy goal of trying to increase access to first-generation, more vulnerable, and poorer students led to pressures not to fail students who were less prepared for college, and to offer online alternatives, often decreasing the quality of course content, so that the value of the completed degree decreased. This may have had the effect of driving the more well-prepared students to higher prestige schools with more valued degrees, leaving the target population even further behind.²⁶

Many of the policy goals in which governors are interested, such as economic growth, are system-targeted in this way. Thus far our focus has been on predicting such variables. *Promoting* important social values such as economic growth requires, in addition, levers that reliably affect the target system as a whole. As Banerjee and Duflo conclude, “Economists (and other experts) seem to have very little useful to say about why some countries grow and others do not. . . . In retrospect, it is always possible to construct a rationale for what happened in each place. But the truth is, we are largely incapable of predicting where growth will happen, and we don’t understand very well why things suddenly fire up.”²⁷ As governors pursue these wider goals, both the predictive and manipulative functions are strained by the complexity of the target system.

Reflexivity (and Using It)

We have already considered reflexivity in detail: a fundamental feature of all policy in human systems is that, while we may know which agents’ decisions are producing what unwelcomed effects, the governor has great difficulty controlling these decisions through institutional rules and incentives, and the interventions themselves are part of a process of reflexive readjustment by agents. Our micro example of seat belt laws (§23.3) is a case in point. We know what causes traffic deaths, but even in what appear to be extraordinarily simple interventions, people can react in unexpected ways—perhaps speeding up, apparently killing more pedestrians and cyclists. The governor can proclaim a directive, but it cannot mandate the desired response. It again bears stressing that there is no greater folly in governance than confusing what the governor directs the agents to do and what, given the governor’s rules and regulations, they do, and what others do in response to both patterns of conformity and nonconformity.

I have stressed the difficulties posed by reflexivity; it is, perhaps, appropriate to note the significant advancements promised by behavioral psychology. Traditional economic advice acknowledged that policies must be “incentive compatible,” but its unrealistic view of what constituted human incentives often undermined this insight.

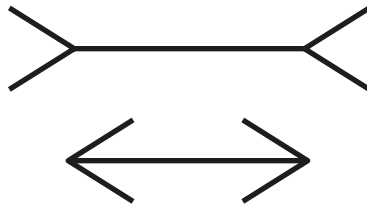
²⁵ Tenner, *Why Things Bite Back*, p. 42.

²⁶ Vedder, “The Flight to Quality in College Admissions.”

²⁷ Banerjee and Duflo, *Poor Economics*, p. 267.

Behavioral psychology offers a more promising route—at least for some cases. The demonstrated effectiveness of “nudges” shows the importance of thinking about reflexivity: nudges operate by predicting what the response to a regulation will be, and reverse-engineering the regulation on this basis.²⁸ Nudges typically take advantage of strong psychological heuristics or cognitive biases; to the extent that we are all subject to these strong biases, we should react in similar and predictable ways. For example, if when designing official forms the governor has good evidence that the reflexive response of most agents will be to take the default option, then the governor can successfully manipulate to secure its goals by assuming this reflexive response in the design of its policy. More generally, designers of policy have employed behavioral psychology to design government assistance programs, health care, retirement plans, and so on.²⁹

However, the limits of nudging and, in general, of behavior psychological design of policies may well be quite severe. Nudges are grounded on a strong cognitive bias that can be co-opted into the governor’s policy goals, and so a high degree of predicted uniformity of response. In many areas, more recent behavioral psychology indicates a diversity of agent types. Whereas earlier research asked, “what do humans do?” (“do they contribute to public goods?”), later research has shown us that in the same population different individuals respond differently. Moreover, many “cognitive biases” may be cultural. For example, it is often assumed that the Müller-Lyer illusion (Display III.9) is an inescapable, strong, universal bias³⁰ (the lines are the same length, but the top seems longer), but cross-cultural research indicates that American undergraduates, who are the most studied group about almost all biases, also have by far the strongest reactions (seeing the top line as about a fifth longer), while the San foragers of the Kalahari see no difference. The rest of the world is in-between.³¹ Thus, as Woodend, Schölmerich, and Denktas note, insofar as biases are cultural, different perspectives may well react differently to nudges, reintroducing a diversity of response.³²



Display III.9. Müller-Lyer illusion.

²⁸ See Thaler and Sunstein, *Nudge*. For doubts about nudging from an Ostrom-informed perspective on self-governance, see Aligica, *Public Entrepreneurship and Self-Governance*, pp. 150ff.

²⁹ For a more theoretical survey of behavioral psychology and some of its policy applications, see DellaVigna, “Psychology and Economics: Evidence from the Field.”

³⁰ “We are all vulnerable [to biases]. And there’s no way to make ourselves bulletproof, as the famous Müller-Lyer optical illusion illustrates.” Tetlock and Gardner, *Superforecasting*, p. 233.

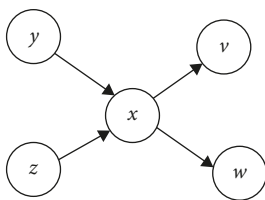
³¹ Henrich et al., “The Weirdest People in the World?,” p. 62.

³² Woodend, Schölmerich, and Denktas, “‘Nudges’ to Prevent Behavioral Risk Factors Associated with Major Depressive Disorder.”

27.4. Multiple Policy Goals: Many-to-One-to-Many Influence

We began this section focusing on single sectoral aims—to manipulate a discrete variable. Within complex systems there are good grounds for concluding that short-to-medium-term predictions of such discrete variables are possible, and in some cases perhaps a system governor can find levers to manipulate them, though revenge effects are always lurking. As Tenner concludes, “What is almost a constant . . . is that the real benefits are not the ones we expected, and the real perils are not the ones we feared.”³³ When the target is a system of variables—such as health, growth, and education—the familiar problems of complexity begin to reappear. Still, if the governor pursued simply one, or a few, sectoral goals, there would be modest prospects for success—it is certainly unfounded to claim that it is impossible. However, as the governor begins to take on additional goals, the problems of complexity bite back. What is the target of one policy is an input variable into another. Not only do we have a many-to-one relation between the influences on W ’s value, but W ’s value is one of many inputs into the value of another target, Y . We seek economic growth, environmental protection, education, healthy lives, secure retirement, employment protection, higher wages, affordable housing, interesting work, leisure, safety, privacy, equity, social mobility, and on and on.

To see the difficulty more clearly, consider a simple simulation where the governor confronts an eight-variable system with four target variables (policy goals). The other four variables are not targeted by policy: suppose for now the problem is cognitive—the governor may not be aware of the impact of the other four variables. This system builds on the simple model in Display III.8, in which each variable (x) is influenced by two others (y, z). But now (Display III.10) we not only consider the many-to-one relation (y and z affect x), but also the one-to-many relation in which x itself affects two other variables (v, w). One policy’s output is, as I have said, an input which affects another policy goal.

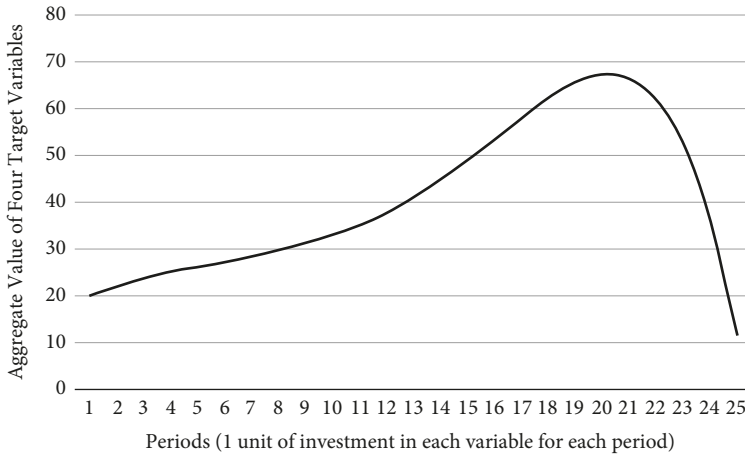


Display III.10. Many-to-one-to-many influence.

In the simulation resulting in Display III.11, each variable in the eight-variable set was influenced by two other variables.³⁴ The influence relation was either positive

³³ Tenner, *Why Things Bite Back*, p. 272. While much of Tenner’s discussion is focused on technological innovations, a good deal also concerns policy interventions.

³⁴ Most variables also influenced two others. In this simulation (the first full run of the model, so it was not the upshot of searching for a certain sort of case) two variables influenced only one other variable, and two variables influenced three other variables. For the full set of equations, see Appendix B.



Display III.11. Four target variables in eight-variable system.

or negative. Given that y influences variable x , it can either be positively monotonic (an increase in y increases x 's value) or negatively monotonic (an increase in y 's value decreases x 's value). The influence factor on x could also be moderate (.2) or weak (.1). (Both the positive/negative and strong/weak relations were determined randomly for each influence). If y 's influence is both positive and moderate, an increase in y will increase x by .2; if z 's influence on x is weak and negative, an increase of 1 unit of z will decrease x by .1. These seem significant but by no means overwhelmingly tight connections between a variable and its two influencers. Because influence direction and strength were determined randomly, a variety of combinations of weak and strong, positive and negative influences were present. In addition, it was supposed that all variables began with a value of 5, and that the governor was successful in every period in adding 1 unit to the value of its four policy targets. Thus, where y and z are members of the eight-variable set, the value of any target variable x , in period i is:

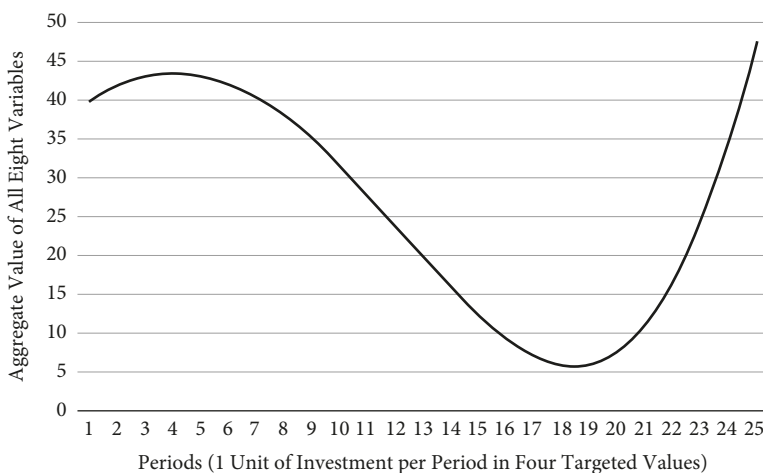
$$\text{EQ. 2} \quad x_i = x_{i-1} + 1 \pm (p \times y_{i-1}) \pm (q \times z_{i-1})$$

where p and q , the influence factors, can be either .1 or .2 (p can equal q), with + or – being whether the variable has a positive or negative effect on x .

Display III.11 gives the aggregate overall effect over the full 25-period run on the four target variables, in which the governor contributes 1 unit to each of the four variables in each of the 25 periods (for a sum total of 100 units investment). While the impact of each variable on the others was strictly linear from period to period, the system is distinctly nonlinear, with a tipping point around period 21, which drove the aggregate values of the target variables sharply downward. Up to period 20 it looked like the policy interventions were succeeding, but the hidden dynamics of the other four variables were about to swamp the policy intervention. A model extrapolating from the data from periods 1–20 would be wildly off by period 25. By the

end, the target variables had lost half their initial value (going from an aggregate of 20 to near 10) despite a 100-unit investment by the governor. And, as we have seen, characteristically of complex systems, the system was predictable up to around period 21—when it surprised us. Thus we see that even a model based on simple linear dynamics of the preceding equation (EQ. 2) can give rise to nonlinear aggregate dynamics;³⁵ with nonlinear micro dynamics, these problems are greatly aggravated.³⁶

Consider another interpretation: suppose all eight of the variables were important to the governor, but only four could be manipulated and thus targeted. In that case, the success of the attempt at manipulation should consider the effect on all eight variables, as in Display III.12. We find that the overall system actually began to improve after the period 21 tipping point; at period 25 it finally exceeded its initial value of 40. Note the more complex nonlinear aggregate relation as more variables are added.



Display III.12. Aggregative values for entire eight-variable system.

This is simply illustrative—a different array of connections will, of course, yield different results. But such a basic model should warn us against our strong intuitions that we can secure multiple policy goals in an increasingly interconnected world. This was a very simple eight-variable world, with each variable being linearly influenced by two other variables, and in which the governor was able to enhance the value of all its targets in each period. A huge number of the actual complexities were bracketed out. Nevertheless, the policy interventions were ultimately swamped by system dynamics. Because we often do not perceive the underlying dynamics (such as the links to the four non-target variables), it is tempting to attribute the dramatic events at period 21 to exogenous shocks or decaying implementation. After all, if the sole concern was the four target variables, things were working reasonably well up to period 20 (there was

³⁵ Beinhocker, "Reflexivity, Complexity, and the Nature of Social Science," p. 333.

³⁶ See Colander and Kupers, *Complexity and the Art of Public Policy*, pp. 117ff.

gain of almost 50 units for 80 units of expenditure). The governor may thus conclude that it controlled the system in periods 1–20—but it really never did.

27.5. Muddling Through

Effective goal pursuit, we have seen, requires both cognitive and manipulative functions. There is evidence that on relatively discrete variables the best forecasters (and even more, groups of diverse good predictors) can have a good cognitive grasp of the direction of the variables up to three years or even five years. Evidence also indicates that by a maximum of five years, predictive power reverts to about chance. In addition, sectoral-goal pursuit requires a manipulative ability. Because of reflexivity and the unknown underlying dynamics of a complex order, there is a world of difference between prediction and manipulation. I have argued that while in various circumstances a single target (or a well-identified target system) can be effectively manipulated over the short-to-medium term, the problems of complexity are apt to undermine multiple goal pursuit in an increasingly interconnected system. As we pursue more goals, the effects of one attempted manipulation rebound onto other targeted variables. This conclusion is resisted by many. Our linear intuitions—that if we can pursue one goal, then pursuing five is only five times more difficult—is deeply misleading. As our simple model shows, the hidden dynamics of even relatively simple systems can be surprising and overwhelming.

Tetlock's second study enhanced the predictive ability of forecasters by allowing them to update as events unfolded.³⁷ Because our models of complex systems are so imperfect, it seems critical that, as in Knight and Johnson's Deweyian account of democracy (§23.1), governance is conceived as a constant reflexive and recursive process, in which the democratic governor reviews the state of the system and its own responses to better adjust to secure its goals. In a classic essay, Charles Lindblom described one such policy method as "muddling" through:

- (1) Selection of value goals and empirical analysis of the needed action are not distinct from one another but are closely intertwined.
- (2) Since means and ends are not distinct, means-end analysis is often inappropriate or limited.
- (3) The test of a "good" policy is typically that various analysts find themselves directly agreeing on a policy (without their agreeing that it is the most appropriate means to an agreed objective).
- (4) Analysis is drastically limited:
 - (i) Important possible outcomes are neglected.
 - (ii) Important alternative potential policies are neglected.
 - (iii) Important affected values are neglected.
- (5) A succession of comparisons greatly reduces or eliminates reliance on theory.³⁸

³⁷ Tetlock and Gardner, *Superforecasting*, pp. 91–2.

³⁸ Lindblom, "The Science of 'Muddling' Through," p. 81.

Rather than relying on well-developed models or comprehensive comparisons of the full option space, the governor makes a series of decisions or, we shall say, “moves” in a severely constrained option space. At each move, it evaluates whether there is some modest adjustment to improve the results given the available alternatives.³⁹ For this to be effective, the actual space must not be wildly different from the constrained space the governor is focusing on. Moreover, the policy space must be relatively stable and the explorer must experience—to use a concept from communication theory—an acceptable signal-to-noise ratio.⁴⁰ Let us consider each in turn.

The Optimization Space

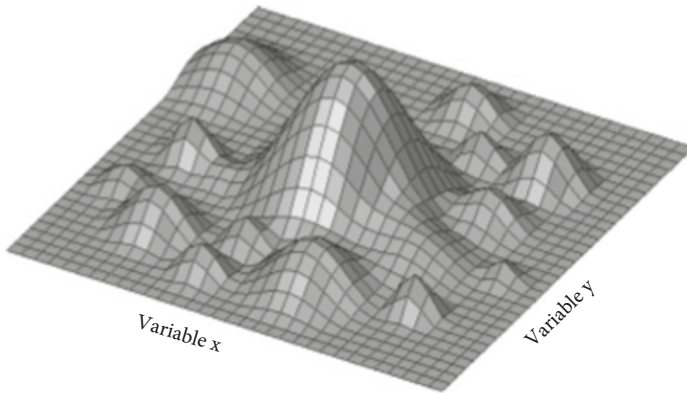
Consider again Display III.8, a many-to-one relation where the “one”—the policy goal—is several closely related variables. This seems a basic case: seldom is the goal (growth, health, security) truly a singular, discrete variable. As we saw with revenge effects, an increase in one component of the system can degrade others (§26.3). An aware policymaker would need an implicit function that takes the scores of each of the component variables and allows her to determine whether an intervention increased the total growth score of the targeted variable-system, given its effects on the component elements. Suppose a policymaker is seeking to promote growth (G), where this is understood as a function of employment (e) and median income (m), so $G = f(e, m)$. Although at one time it was thought that these two are simply monotonically related (one increases with the other), recent experience has uncovered a more complex relation, whereby high levels of employment may not spur increases in median income, or at least may only very weakly do so. And perhaps very high increases in median income could only result through transfer programs, which would reduce the incentive to work. So G is a cumulative, but seldom a simple aggregative measure. Taking a cue from Lindblom’s point (*i*), the function $G = f(e, m)$ will be both empirical and normative. Empirically, it will seek to capture interaction effects (such as revenge effects) between m and e : for any given values of e and m , the policymaker wants to know how they interact and so affect each other. But the function will also express the policymaker’s evaluation of, for every level of e and m , how much growth (G) is being generated. In that sense it will also be normative: specifying G manifests a judgment about the relative importance of e and m .

Thus far we have modeled this many-to-one (system) relation in terms of multiple regression or curve-fitting models, but neither of these seems appropriate to muddling through: they suppose a pretty robust theoretical orientation. A model more in tune with muddling through is a “rugged landscape” search model as in Display III.13. This sort of model is a common method to analyze optimizing (or at least getting better) in complex systems through localized searches.⁴¹ The x and y axes are

³⁹ Ibid., p. 84.

⁴⁰ For an analysis of complex system in terms of networks and signals, see Holland, *Signals and Boundaries*.

⁴¹ See the *Tyranny of the Ideal*, chap. 2. The idea of a “rugged landscape” is not a mere “metaphor”—at least no more metaphorical than “game theory”—it is the result of a mathematical analysis of complex systems. See my “Political Philosophy as the Study of Complex Normative Systems.” Muldoon’s model in section 14.3 is of this sort.



Display III.13. A moderately complex two-variable policy landscape.

scores on two policy variables, the z axis the cumulative effect. The terrain depicted is characteristic of an NK system with a low number of interconnections between the dimensions (goals), so it is by no means highly complex.⁴² That we are assuming such a landscape is more or less stipulated by Lindblom's clauses (4.i) and (4.iii): we are simply ignoring the effects of our decisions on many values to simplify our search problem, and so effectively reducing the complexity of the problem for our purposes.

A muddling-through learning strategy can get the governor up a gradient; at each point it reflexively re-evaluates its options and pursues that which yields a somewhat better result. The white line illustrates a muddling-through strategy that gets stuck early on at a poor result: from this poor peak, all muddles seeking a better result are frustrated—all moves make things worse. But in this low-dimensional (not very complex) landscape, it is not all that difficult to muddle one's way through to the best solution by simply looking for an improvement at each step. A basic learning strategy of always looking for an improvement will help the governor climb a gradient until it gets to an "optimum" from which every change decreases the aggregate (z) value of variables x and y . More sophisticated strategies allow for searching that sometimes accepts worse outcomes ("going downhill") in the hopes of finding a bigger mountain to climb; for example, early on in the searches it may tolerate going quite far down in its search for a higher peak, but as the search continues (and presumably it has explored a lot of the territory) it may be less tolerant of such "mistakes" and becomes more conservative.⁴³

The success of such "trial and error" searches depends on several conditions:

- (i) Most obviously the optimization landscape cannot be too complex. We initially accepted Lindblom's supposition that we simply ignore other values, and that this had the practical upshot of making the problem less complex. But the lesson

⁴² When $K = 0$ a "Mount Fuji" landscape results, where, from any point, there is a steady gradient to the top. As we have seen, when $K = N - 1$, the terrain is chaotic, with many sharp peaks and valleys, in which the peaks tend to be low.

⁴³ See Page, *The Difference*, pp. 64ff.

from our basic simulation (Display III.11) holds: though the governor ignored all but four values in the eight-variable system, the other four were still in play, creating an unwelcomed drop of the target variables into a deep valley just when the governor thought it was heading uphill. In this sense, the other values refuse to be ignored. If the actual landscape is too complex (if there are many interconnections among many variables), there are apt to be numerous peaks and valleys in the targeted variables. In such highly complex landscapes, trial and error is very likely to quickly hit a poor local optimum: a mediocre solution may thus be “locked-in” by the agency.⁴⁴

- (ii) If the governor is to muddle its way through to better polices by searching the optimization landscape, the landscape cannot change too swiftly. As I have stressed, education policy, economic policy, social welfare policy, and national defense policy all interact. When we disaggregate these, we will find that each area presents its own rugged landscape: education departments seek to “muddle through” on pursuing their own aims, while economic units explore better policies through their own searches on their own “landscapes.” Because the effects of policies are interconnected, as one department makes decisions (and, so, moves to a new spot on its landscape), it affects the landscape of others: having changed economic policy, what works as an effective education policy is now different. We have already encountered these “dancing landscapes” (§17.5).⁴⁵ Once two systems are coupled in this way, optimization dynamics become much more complex; neither system can be strictly modeled as having local optima that serve as steady attractors for searches. Indeed, it can be difficult to model such systems in terms of optimization at all.⁴⁶ We have seen (§17.5) that under some conditions uniting high K (highly interconnected) systems can *decrease* complexity; however, if the two systems are tightly bound, complexity in each will tend to increase.⁴⁷ Under these conditions, coupling rugged landscapes can lead to drastic fluctuations in each landscape. The landscape may well become too changeable to allow effective muddling through learning. One policymaker can seek to ignore the effects of its policy on another department, but as that department adjusts, it changes the very problem the first policymaker thought it had made progress on.
- (iii) And, of course, we must add to the problem of dancing landscapes the ever-present fact of the reflexivity of social systems. As I have rather harped on this point, consider Alex Rosenberg’s analysis on the changing terrain of pursuit of justice:

Human action ... is highly reflexive. Among humans, choices are almost always to some extent strategic, not parametric. People choose their behaviors to line up with

⁴⁴ Of course this is not to say that there are no responses, seeking to develop more complex strategies that avoid such traps. But it is exceedingly difficult to successfully search such landscapes. See Levinthal and Warglien, “Landscape Design: Designing for Local Action in Complex Worlds.”

⁴⁵ Page, *Diversity and Complexity*, p. 94. See further my “The Evolution, Evaluation and Reform of Social Morality.”

⁴⁶ Kauffman, *The Origins of Order*, p. 238.

⁴⁷ *Ibid.*, p. 252.

or exploit other people's strategies.... Since which individual and group strategies are chosen often depends on which strategies are already in play, the space of human institutions and behaviors is a constantly shifting landscape in which there are few regularities that obtain long enough for policy planning to actually exploit in the design of institutions.... The terrain of justice is a continually changing surface... hill-tops grow and shrink continually, even move, fission, fuse, and most complicating of all, continually create new features—valleys, gullies, troughs, berms, bluffs, hills, mountains, cliffs, and so on, as they come and go.⁴⁸

Governors cannot muddle through to reasonably good results in such a continually changing landscape. Even if an agency increases the values in its remit at every decision node, the cumulative effect is too likely to be wandering around an ill-defined space.

The point of (i)–(iii) is not that reasonably effective “muddling through” trial-and-error searching of a complex policy terrain is impossible, but that the problems of degree of complexity, coupled landscapes, and reflexivity must be greatly mitigated for the searches to yield reasonably cumulative results.⁴⁹ If they are not, muddling through looks simply muddled. It seems rather dubious that they can be significantly mitigated for sectoral goals encompassing a large and diverse system.

Noise

Implicit in the landscape analyses is that the searchers have reasonably reliable information as to when a search is moving up or down a value gradient.⁵⁰ If an explorer is blind to evidence of whether she has gone up or down, her movements will be random. The explorer thus needs some reliable signals of the results of her action. Importantly, it may be best if the information is not perfectly accurate. Some noise may allow our explorer to traverse gullies: not being able to immediately determine that they have started to move down, the policymakers might have proceeded for a while and hit the upward gradient, and so did not get stuck at a local optimum (a low local peak).⁵¹ But if there is too much noise, the search becomes random. At the macro and higher meso levels, it is extraordinarily hard to determine what changes in the system were produced by one's own actions (what worked better because of a department's innovation) and which macro changes were the result of exogenous events. Sometimes innovators cope with this by adopting a myopic assumption that unless it is obvious that they caused an event (there was clear and direct causal sequence), a change was exogenous to their actions.⁵² But as we saw in our simple simulation, often the effects

⁴⁸ Rosenberg, “On the Very Idea of Ideal Theory in Political Philosophy,” pp. 62–4. Paragraph breaks deleted.

⁴⁹ This is the point of Levinthal and Warglien, “Landscape Design: Designing for Local Action in Complex Worlds.”

⁵⁰ See Page, *The Difference*, p. 160.

⁵¹ Miller and Page, *Complex Adaptive Systems*, p. 128.

⁵² Or, “like players in a game of Mikado (or pick-up sticks) policy makers have chosen to focus exclusively on those sticks that you can pick up without influencing the others. The problem is that in real-life social systems, most of the sticks hang together, so you need to find entirely new strategies to deal with clusters of interconnected sticks.” Colander and Kupers, *Complexity and the Art of Public Policy*, p. 14.

are at least partially the results of one's own interventions, and one can believe one is successfully muddling up a gradient as one is about to fall into canyon. There is a great difference between muddling through and finding better solutions, and thinking you are muddling through when you are actually mucking around.

27.6. Meso or Micro Learning-Based Governance?

While the conditions for learning at the macro or higher meso level are not auspicious, as we move "lower" into the system, the possibilities of such learning-based governance are enhanced. The critical question is at what level of complexity learning-based governance becomes a reliable way to secure policy goals. Perhaps the most detailed and optimistic analysis is Sabel and Zeitlin's "experimentalist governance" model of the European Union. Their analysis focuses on EU governance for a number of policy areas, including "telecommunications, energy, drug authorisation, occupational health and safety, employment promotion, social inclusion, pensions, health care, environmental protection, food safety, maritime safety, financial services, competition policy, state aid, anti-discrimination policy and fundamental rights."⁵³ The European Union, they argue, has developed a "new form of governance" of "directly deliberative polyarchy" which, they claim, allows it to "respond effectively to conditions of persistent diversity, complexity, and volatility."⁵⁴ On this model:⁵⁵

- Overall policy goals are jointly formulated by EU institutions and member states;
- Lower-level agencies are given the freedom to advance these ends as they see fit and are empowered to propose changes in them;
- Lower-level units must report back their decisions and performance to "peer reviews" according to agreed-upon indicators, and comparison with other agencies pursuing similar ends.
- The results of lower-level efforts are recursively considered by the higher-level agencies, which then can revise their goals given implementation data.
- Deliberative participation by diverse interested parties is part of this ongoing review process.

As they see it,

These governance processes may also be considered a form of "directly deliberative polyarchy" (DDP). They are deliberative because they use argument to disentrench settled practices and open for reconsideration the definitions of group, institutional, and even national interest associated with them. They are directly deliberative because they use the concrete experience of actors' different reactions to current problems to generate novel possibilities for consideration rather than buffering decision-makers from mundane experience, the better to elicit their principled,

⁵³ Sabel and Zeitlin, "Learning from Difference," p. 271.

⁵⁴ *Ibid.*, pp. 276, 323.

⁵⁵ *Ibid.*, pp. 273–4.

disinterested response to abstractly posed problems. And these governance processes are polyarchic because, in the absence of a central, final decider, their constituent units must learn from, discipline, and set goals for one another.⁵⁶

While by no means dismissing the insights from Sabel and Zeitlin's studies, many of the worries encountered in this section's discussion of sectoral goals re-emerge. Most obviously, their proposed model of governance is firmly rooted in the aspiration to secure many-to-one control, where the target variable is a system of related variables. We have looked extensively at the problems raised by numerous agencies treating their behavior simply as ways to secure their own goals, rather than also being inputs into those of others (§27.4). Let us bracket that now well-explored problem, and focus simply on the agency's many-to-one project.

Sabel and Zeitlin, we have seen, understand the Deweyian recursive system of governance as a response to diversity and complexity.⁵⁷ For Lindblom, too, muddling through without reliance on powerful theories is a way of coping with complexity.⁵⁸ Drawing on the rugged landscape model of section 27.5, learning-based approaches to complex problems can be understood as using search heuristics to explore a complex landscape. We also uncovered desiderata for successful learning: the landscape cannot be too complex, cannot change too quickly, and there cannot be too much noise about the effects of interventions. This suggests that Sabel and Zeitlin's proposal, with its strong dependence on recursive learning, is most likely to be successful in systems that are not tightly coupled: there is no great value in peer review of past policies if we cannot reliably identify what were the endogenous and exogenous effects of past protocols. Our intuitive confidence that we can do this is often misplaced: we focus on easily observed effects and discount more remote consequences. Recall the lesson from Parts I and II: the functionings of our norms, practices, and institutions are often "*causally opaque*—an individual cannot readily infer their functions, interrelationships, or importance," and so "intuitions and personal experiences can lead one astray."⁵⁹

These considerations point to the greater effectiveness of muddling through by iterated learning when systems are less coupled and reflexivity is minimized: when both conditions are met, it is easier to gauge the effects of past actions and to climb a steady gradient.⁶⁰ Interestingly, these are just the sort of systems on which the Ostroms's polycentrism focused: smaller-scale strategic dilemmas. Like Sabel and Zeitlin, and Knight and Johnson, the Ostroms focused on devising institutional structures that could learn through contestation and deliberation, with constant feedbacks concerning the effects of not only the first-level rules, but also the second-level rules that structured responses to strategic dilemmas. And because they focused on common

⁵⁶ Sabel and Zeitlin, "Experimentalist Governance," p. 170.

⁵⁷ Knight and Johnson view recursive democracy as a response to diversity and Hayekian information problems. *The Priority of Democracy*, chap. 3.

⁵⁸ Lindblom, "Muddling Through," p. 72.

⁵⁹ Henrich, *The Secret of Our Success*, pp. 99–100, emphasis in original.

⁶⁰ "The most successful of these arrangements combine the advantages of decentralised local experimentation with those of centralised coordination, and so blur the distinction between forms of governance often held to have incompatible virtues." Sabel and Zeitlin, "Learning from Difference," p. 275.

pool resources problems, the reflexivity of responses was greatly dampened. To be sure, Elinor Ostrom stressed that the scope of polycentric government was by no means restricted to micro (and, we might say, lower meso) levels, but the brunt of the Bloomington School's research was indeed at this level. Still, because focusing on strategic dilemmas blunts reflexivity—and thus one of our two confounding influences on steady learning is blunted—if we can rise to “experimental” learning at higher levels, it looks as if it will usually be on solving strategic dilemmas.

§28 Self-Governance from the Bottom Up

Simplifying the Problems of Governance

28.1. Agent-Based Models

By this juncture the characteristic features of complex self-organized systems should be familiar: agents are pursuing their own (diverse) ends, they reflexively respond to the actions of others, their interactions are structured by a set of rules, and the outcome often depends on who interacts with whom (for example, what others are attracted to the new niches one has created or, more prosaically, who is one's neighbor). We also do not assume that people are hyper-rational calculators: they tend to make “adaptive” choices—ones that they see as better for them given the relatively restricted set of options they are considering. But they also update and change their “strategies,” devising new reflexive responses. And, because of all this, self-organized complex systems tend to be path-dependent: where the system ends up often depends on who first met whom, and what changes reverberated from there.

The most promising way to understand such systems is through the general family of “agent-based models.”¹ The crux of such models is to create a world of agents with beliefs, ends, networks of relations, and rules of interactions, and simulate how this mini social world can develop. In some cases, such as our model in section 20, there might be a unique result, but in many, each run of the model will yield a different outcome. But many runs of the model over many iterations may give us—as Hayek might say—an understanding of which patterns are more likely to emerge. In this sense, the models are predictive, though it should always be remembered that the prediction depends on the assumptions about agents, rules, and networks made in the model itself. These models can be most enlightening, but as populations increase, problems of computational complexity arise.

¹ There is a plethora of types of models available to a social science that is aware of complexity. See Page, *The Model Thinker*. For an accessible governance-based discussion, see Lane, the *Complexity of Self-Government: Politics from the Bottom-up*.

An interesting example is a recent agent-based model by Gressman and Peck on the growth of COVID-19 in a large research university.² Their model is populated by 22,500 agents in a university. Every run of the model has 20,000 students and 2,500 instructors who interacted daily for 100 days. Every day, simulated individuals experience contacts with some others; contacts are scored to vary with degree of contact (friends in a class, say, have a higher contact score than strangers in the same class), and residential contacts are distinguished from class-based contacts. Contacts, of course, are means of Covid transmission. The model also includes assumptions about the disease and responses:

- *Illness Testing*: A predetermined fraction of the population is randomly selected for testing. The default value is set at 3%, which means that members of the community will be tested approximately once per month. Anyone who was tagged via contact tracing on the previous day is also tested.
- *Quarantining*: Anyone from the Illness Testing step whose test results were positive and any symptomatic infected individuals who develop symptoms on this day are quarantined immediately (if not already quarantined). Additionally, such individuals are tagged to have their recent contacts traced.
- *Status Updates*: Individuals who have been quarantined for 14 days are released....
- *Contact Tracing*: Every individual who was tagged earlier in the day (due to testing positive or developing symptoms) has their contacts traced.... These contacts are tagged for testing tomorrow and are quarantined immediately.
- *Infection Transmission*: Non-quarantined susceptible individuals who contact non-quarantined infected individuals become infected themselves with a probability that depends on the infection state of the infected....
- *Outside Transmission*: On any particular day, there is a 25% chance that one non-quarantined susceptible individual becomes spontaneously infected due to presumed transmission from non-university contact....
- We assume that 75% of those infected are asymptomatic and that asymptomatic individuals are half as infectious as symptomatic counterparts.
- There is a global parameter to account for a variety of nonpharmacological interventions (e.g., mask wearing). The default value is chosen so that transmission probabilities are reduced to 50% of the values identified above. This decreased transmission is included to reflect the expected effects of universal mask-wearing policies....³

Other parameters include the reliability of the testing, mean class size, etc. Various levels of university interventions were also considered: the standard intervention “consists of the combination of quarantine and contact tracing, universal mask-wearing, daily randomized testing of 3% of the university community, and transitioning all classes with 30 or more students to online-only interaction.”⁴ Thus, in a

² Gressman and Peck, “Simulating COVID-19 in a University Environment.”

³ Ibid., pp. 3–45. The points have been reformatted, and citations omitted.

⁴ Ibid., p. 9.

sense, a virtual university has been created to understand the effects of certain preventive interventions on the self-organized complexity of this microworld. Gressman and Peck find support for the effectiveness of interventions: “Although implementing the standard intervention is costly, it is also crucial for controlling disease outbreaks. In the absence of any intervention, all scenarios end with effectively all susceptible community members developing COVID-19 by the end of the semester, with peak infection rates reached between 20 and 40 days into the semester. . . . In contrast, the standard intervention avoids the epidemic tipping point altogether and keeps cumulative infections below 66 in more than 95% of simulations.”⁵ Gressman and Peck also disaggregated the efficacy of the interventions, finding transition to online of classes over 30 highly effective, wearing masks moderately effective, while random testing and contact tracing had the lowest efficacy.

All agent-based simulations have limits imposed by assumptions made (and not made) about parameters,⁶ rules of interaction, and the structure of network interactions. Now in theoretical models that we develop to better understand some dynamic,⁷ a model can have just a few parameters to highlight the dynamic in question. As Gressman and Peck’s analysis makes clear, as the model seeks to have direct policy implications, the numbers of relevant parameters increases, which in turn vastly increases sources of error. This is by no means to say that agent-based models are of no use in developing policy, but we can see that the complexities can mushroom, even in a relatively small-scale environment. To be persuasive the parameters must be based on rich empirical data; aggregating a number of different models would be necessary before we rely on them for serious policy advice (see further §30.1). Nevertheless, agent-based models are powerful and appropriate tools for understanding self-organized complexity and the effects of interventions. As guides to policy in highly diverse and deeply interconnected systems where relatively precise predictions are required, they are manifestly most useful in smaller-scale contexts.

28.2. True Experiments, the Microlevel, and Social Improvement

Recall that Sabel and Zeitlin deemed their proposal “experimentalist governance” (§27.6). While this is perfectly acceptable usage (as is Mill’s “experiments in living”), in an important way it obscures just how far short their procedure falls of the scientific ideal of an experiment. Because contextual learning does not seek to identify the relevant causal factors yielding an outcome, its results are not replicable. We should not expect them to transfer, since “in many domains the causes that operate shift frequently and unpredictably, from locale to locale and from time to time, as economists

⁵ Ibid., p. 10.

⁶ E.g., Gressman and Peck did not account for “the fact that compliance with mask-wearing policies will likely be diminished in residential settings and social settings; our modeling leads us to believe that this will lead to modest increases in infection and quarantine rates.” Ibid., p. 15.

⁷ For example, the famous Schelling model of segregation that explains segregated residential patterns with very few assumptions. Schelling, *Micromotives and Macrobehavior*, chap. 4.

from John Stuart Mill to British econometrician David Hendry have argued.”⁸ If the Swedes have muddled through to an acceptable welfare policy, there is no good reason to think that the Americans can take over this policy with similar results. The Swedish policy is embedded in a distinct complex system, with different connections and norms that have evolved along their own unique path. If we wish to know *why* a policy works and *where* it will also work, Nancy Cartwright and Jeremy Hardy insist that we need to keep three points in mind: “1. The causal principles that underwrite policy prediction are not universal. 2. Few causes work on their own; causal factors work together in teams. 3. There are generally a number of distinct teams at work in any situation, each making its own contribution to the effect.”⁹

Because of these facts, even the “gold standard” of social experimentation—the randomized control trial—cannot tell you much in a single case. At best, such a trial will only tell us “that a policy worked there, where the trial was carried out, in that population.”¹⁰ To know what interventions work in different environments, we need to know the relevant causal factors, and whether they work in that environment. As Cartwright and Hardie put it, the basic claims we need to establish are:

- (i) The policy worked there (i.e., it played a positive causal role in the causal principles that hold there, and the support factors necessary for it to play this positive role there were present for at least some individuals there).
- (ii) The policy can play the same causal role here as there.
- (iii) The support factors necessary for the policy to play a positive causal role here are in place for at least some individuals here post-implementation.¹¹

Our ability to ground these claims is enhanced by conducting a number of trials in different systems to understand the causal factors and the necessary background (supporting) conditions. Yet, while “[l]ots of positive RCT results are a good indicator that the policy plays the same causal role widely enough to reach to you[,] . . . you are always betting on a hidden premise: that the studies vary across enough different kinds of circumstances to generalize—and indeed across just the ones that matter for your situation. That bet is always a little dicey.”¹² And it is especially dicey in highly complex systems, where connections abound, and causal teams can appear and disappear. As we zoom down to smaller systems, these problems often become more manageable. I have argued that at the higher-meso level, many of the influences on behavior typically derive from forces elsewhere in the system. The meso system is itself large, and has many features that are woven into other subsystems. However, as we get down to micro levels—say, a neighborhood—it becomes significantly simpler and many of its features are endogenous. Whether there is crime is no doubt part of the national and urban areas in which it exists; whether there is a neighborhood watch and adequate street monitoring are a result of endogenous forces. It is always a matter

⁸ Cartwright and Hardie, *Evidence-Based Policy*, p. 42 (citations deleted). See also Beinhocker, “Reflexivity, Complexity, and the Nature of Social Science,” pp. 335–6.

⁹ Cartwright and Hardie, *Evidence-Based Policy*, p. 52.

¹⁰ *Ibid.*, p. ix.

¹¹ *Ibid.*, p. 54.

¹² *Ibid.*, p. 127.

of mixed influences, but numerous features of life in a hospital, a school, a neighborhood, or a university are apt to be the result of endogenous variables and decisions. To the extent that this is so, they can be usefully studied through randomized control trials. Additionally, because there are so many micro systems—such as towns and villages, hospitals, schools, and neighborhoods—a sufficient number of trials can be conducted so that we can establish not just what worked here, but begin to get an idea what might work there.

At this point, readers interested in social policy might moan—we are interested in shaping our society, not fixing our schools and hospitals! As Banerjee and Duflo observe, investigations of poverty far too often have been “fixated” on the “big questions”: “What is the ultimate cause of poverty? How much faith should we place in free markets? Is democracy good for the poor? Does foreign aid have a role to play?”¹³ Yet, as their work on randomized control trials has famously shown, real progress is secured on big problems, such as poverty, failure of education, inadequate health care and disease, by “the accumulation of a set of small steps, each well thought out, carefully tested, and judiciously implemented.”¹⁴ Randomized control trials have provided the basis for successful inoculation policies, nutrition policies, sanitation improvements, enhancing women’s participation in the labor market, overcoming gender inequality by empowering women to invest and to improve their access to education.¹⁵ Indeed, it was not until some economists stopped believing that they could control larger systems and so began to focus on the micro that they have begun to even understand what poverty is, and the economic lives of the poor.¹⁶ We find that the global poor are not simply people without income, education, or enough to eat. They tend to have distinctive economic lives, though these also vary from place to place. Generally, for example, even the very poor see themselves as having significant economic choices, and are not apt to spend additional income on increased calories, though they often opt for higher-quality food, as well as expenditures on wedding and festivals. Entertainment may be of special importance to the very poor, who spend so much of their lives hungry. “Poverty is not simply a shortfall of money. The constant, day-to-day hard choices associated with poverty in effect ‘tax’ an individual’s psychological and social resources.”¹⁷

The fascination with grand attempts to control the system, or improve “Health” and “Education” or abolish “Poverty” by a central controller, not only seldom succeed, but have allowed policymakers to simplify and homogenize target populations such as the poor and the unhealthy. Perhaps this is the greatest paradox of all: by focusing on the most complex level, policymakers oversimplified the less complex levels. If the poor are those who don’t have enough income, then provide them with income and there

¹³ Banerjee and Duflo, *Poor Economics*, p. 3.

¹⁴ Ibid.

¹⁵ See Banerjee and Duflo, *Poor Economics*. See also World Development Report 2012: *Gender Equality and Development*.

¹⁶ Banerjee and Duflo identify “the extremely poor as those living in households where the consumption per capita is less than \$1.08 per person per day, as well as the merely ‘poor’ defined as those who live under \$2.16 a day using 1993 purchasing power parity (PPP) as benchmark.” “The Economic Lives of the Poor,” p. 143.

¹⁷ World Development Report 2015: *Mind, Society, and Behavior*, p. 80.

will be no poverty. By proceeding to build policy up from the micro levels, the first step was to understand that the micro level itself has a certain complexity—the lives of the poor are systems of attitudes, beliefs, options, and norms, as well as resources. Having understood the poor in their complexity and variety, the hard work of determining how to inoculate children, curb open defecation, improve diet, and educate girls could begin with true experimental trials. To abandon the illusion of systematic or meso goal pursuit does not mean that the democratic governor must abandon social improvement, but that it must rid itself of delusions of grandeur and commence the job of finding out what can really work to make people's lives go better. This may not be as exciting as the grand pursuit of the ideally just society—or socialism or libertarianism in our time—but exciting political projects rarely actually improve the lives of citizens.

28.3 Norm Change: Public Justification vs. the Impatient Moralist

I have repeatedly drawn on the development literature, as it has been a comprehensive, careful, and self-critical project seeking social change to secure valued goals. As I have stressed, it has manifested a tendency to move from early focus on macro goals and grand theories through meso goals with increasing focus on micro situations where interventions can be rigorously studied and evaluated. The best development work is grounded in behavioral psychology and economics, with extensive efforts to employ randomized control trials. Another dimension of this movement to the microlevel has been “Community-Driven Development.” Rather than the governor being simply an “external change agent” seeking to guide the lives of the poor toward exogenous (to the community) goals, Community-Driven Development seeks to work with communities, both in developing and refining goals, and working toward their implementation. Recent research has found that successful Community-Driven Development initiatives require more than “change agents” seeking to convince community members of the benefits of participation: actual changes in the social networks and the patterns of interaction appear important in inducing changes in cooperation, securing, for example, health and education objectives.¹⁸

Thus we come back full circle—the importance of self-organized norm networks. Social change often requires a transformation in the expectations that community members have of each other.¹⁹ Gender equality, sanitation, defecation, female genital cutting, nursing infants, and family planning are just some of the social outcomes influenced through changes in norms.²⁰ And by no means are such interventions unique to developing countries. Colander and Kupers's analysis of public policy in rich complex societies puts great stress on the importance of policy to provide space

¹⁸ World Development Report 2015: *Mind, Society, and Behavior*, pp. 50–1.

¹⁹ See Bicchieri, *Norms in the Wild*, chap. 3. I am grateful to Tauhidur Rahman for discussions about his research on some of these matters.

²⁰ For different cases, see World Development Report 2012: *Gender Equality and Development*; World Development Report 2015: *Mind, Society, and Behavior*; Bicchieri, *Norms in the Wild*; Banerjee and Duflo, *Poor Economics*.

for the development of new norms—because of complexity, they argue, “as part of its economic policy, a society should have a norms policy, through which institutions are developed to better allow people to express their collective choice about what norms and tastes should be encouraged and discouraged by society, and that such a policy should be integrated into the institutional structure of society.”²¹

A successful example was an intervention in the United States to reduce bullying in school. Highly connected students and “highly salient” clique leaders participated in a program designed to broadcast students’ experiences with and reactions to harassment and to facilitate public discussion on the issue. The “social referents” wrote and read aloud essays about harassment, performed skits demonstrating the emotional effects of bullying, and sold wristbands signaling the wearers’ commitment to reducing harassment. Changing the behavior of social referents changed peers’ perceptions of schools’ collective norms as well as actual harassment behavior through the mechanism of “everyday interaction.” A key to success in many interventions is to identify the group or social network within which a relevant norm is enforced. Is it the family, the friendship group, the peer group, the neighborhood, or the entire community?²²

Successful norm-based interventions require some approximation to “public justification”: the change must be discussed, debated, and widely accepted by the relevant public for the development goals to be secured. Now here it may well seem that such bottom-up social change is too conservative: perhaps the progressive governor should employ coercion to produce just social change. The case for such action seems, at least on first look, compelling. If practices such as female genital cutting and child marriage²³ are violations of basic human rights to bodily integrity and autonomy, then surely a human rights perspective will insist that these practices be immediately terminated. To maintain that norm change must proceed through justification, appealing to the reasons acknowledged or endorsed by those in the relevant networks can only slow down, and may perhaps even derail, the establishment of human rights. As Bicchieri acknowledges, “There is a certain sluggishness in personal normative beliefs of the ‘moral’ kind, as opposed to prudential ones.”²⁴ Nevertheless, I believe that the methods developed by Bicchieri, stressing the fundamental importance of people in a network recognizing reasons to change their norms and seeking norms that meet the condition of Robust Justification (§§9.3, 21.1), have three great advantages over what we might call the “impatient moralist” approach.

²¹ Colander and Kupers, *Complexity and the Art of Public Policy*, p. 183 (see generally pp. 182ff.). See also Sunstein, *How Change Happens*, Part I.

²² World Development Report 2015: *Mind, Society, and Behavior*, p. 54, reporting on the research of Paluck and Shepherd, “The Saliency of Social Referents.”

²³ We must be careful; neither of these is a uniform practice throughout different cultures. What constitutes female genital cutting or child marriage can differ greatly from culture to culture. The World Health Organization distinguishes four types of female genital cutting. See Hehir, *The Emperor’s New Clothes?*

²⁴ Bicchieri, *Norms in the Wild*, p. 127.

- (i) The most obvious—and to many the most compelling—reason to reject impatient moralism is that it tends to fail to secure the very human rights it values. “[P]lenty of prior experience (especially in developing countries),” Bicchieri notes, “tells us that change initiated in a top-down fashion, typically in the form of legislative interventions, seldom works.”²⁵ We have surveyed some of the evidence in section 22.2. When activists enlist coercive national or international institutions to force people into respecting human rights, the typical result is that the subject population ignores the legal requirements when external enforcement agents are not on the scene, leading to increased disrespect for the law as a mode of social coordination. As the social experiments of Mockus indicate (§22.2), even when law is employed, it must not stray too far from social norms, and the law itself seems to depend on a social norm of obedience.

An important reason for this lack of efficacy is that a simple prohibition of an ongoing practice often takes no cognizance of the need to replace the norm with alternative ways of acting for the members of the network. Take, for example, a law that forbids the marriage of fifteen-year-old girls. If the only option open to girls is to become wives and mothers, and if virginity is seen by potential husbands as an absolute qualification for marriage, remaining unmarried in late adolescence may severely harm a girl's life prospects.²⁶ To effectively undermine the norm, an increased opportunity set for women is typically required; increased employment opportunities for women can provide ways for girls who are educated into their late teens to contribute benefits to the household that would be valued by potential husbands, helping to change the idea of a “good wife.”²⁷ In a network-based public justification—one which is highly sensitive to the personal normative beliefs of the participants (§§9, 20)—these opportunity costs and potential trade-offs would very likely come to the fore. Any reflective network-based decision to abandon a norm (or adopt a new norm requiring the education of daughters) would almost certainly have taken these factors into account, helping to render the collective decision to change efficacious. In contrast, a moralist's dictum that the people in this network simply *must be made* to educate their daughters and not allow early marriage is almost sure to miss these complex issues regarding the other norms in which this norm is embedded and what opportunities might facilitate adoption of the new behavior by parents.

- (ii) This brings us to the second point: when the impatient moralist demands that those in network must (now!) be made to abandon a norm, she very likely fails to understand the function that the norm is serving in the social life of the network, how that function could be better secured by alternative norms, or how it is embedded in the scripts and schemas shared by members.²⁸ The social rules that we live by are not merely specifications of appropriate behavior (conditional on the behavior of others); they typically serve important and diverse

²⁵ Ibid., p. 144.

²⁶ I have greatly benefited from discussions with Gerry Mackie about these matters.

²⁷ Bicchieri, *Norms in the Wild*, pp. 131ff.

²⁸ Bicchieri and McNally, “Screaming Sirens—Schemata, Scripts, and Social Norms.”

functions, helping communities to cope with social dilemmas, tragedy of the commons problems, and, more generally, securing coordinated expectations about justice and fairness.²⁹ In Hayek's terms, a great deal of local knowledge is necessary to even begin to competently evaluate a social norm. Even those in the community may be blind to many of its functions, but external agents are apt to be deeply ignorant. Such knowledge can only be obtained by drawing on the information of the participants in the network. Those with general moral commitments, advancing broad condemnation based on abstract human rights judgments, are extraordinarily liable to miss much that is important in coming to a well-considered conclusion about the justification of a norm. When (as has been done) those who would change a practice of another community depict it as "barbaric"—the practice of barbarians, i.e., those who are "savage, heathen, brutes, or beasts"—we can infer that they may have well failed to fully understand it, and the community they would change.

This is by no means to say that external criticisms cannot be helpful to the participants in better understanding their own norms and their failures. Gender presuppositions often exclude women from the set of recognized evaluators of norms—a state of affairs that women themselves might endorse. In this case, in my terms, the norm might meet the minimum notion of justification implicit in the Justification Effect (§9.3). External criticisms can help challenge the status quo.³⁰ Externally endorsed interventions such as the Tostan Community Empowerment Program can assist in bringing about a more inclusive evaluation of the community's rules—helping to bring the evaluation closer to public reasoning and Robust Justification.

As Bicchieri has stressed, social norms do not merely guide behavior—they provide a system of normative expectations. When one fails to meet these expectations, others in the group are liable to gossip, criticize, and otherwise sanction such failures (§9.4). When one violates a social norm, one expects, as it were, that others will call one out for it. In a practice of accountability, when another calls me out for violation, she is holding me accountable for my violation. I have modeled this in terms of a challenge and response dynamic (§10.3): another challenges my violation, and if I am unable to offer a response, I am held to be at fault. If I endorse the norm, when called out for a violation I am apt to feel guilt and remorse—I may apologize and will seek to alter my behavior. This is a fundamental way in which social norms order the life of a community, and in so doing change the character of the social relations of the members of the norm network. We see each other not simply as objects of control, but as participants in a structure of norms that define relations of accountability, reason-giving, and justification.

Suppose our impatient moralist governor, forsaking public justification, simply imposes a moral judgment on some group or community—and let us even grant that she is morally correct. And suppose even further that it is successful: it has managed to structure the system of incentives that includes punishments for those engaging in the prohibited behavior. But although we have stipulated that the moralist

²⁹ Bicchieri, *The Grammar of Society*, chap. 1.

³⁰ Sen, *The Idea of Justice*, esp. chaps. 6–7.

governor can control behavior, she does not succeed in establishing a practice of accountability. When another calls one out for engaging in the behavior, one will not feel guilty, since one's personal normative beliefs do not endorse the prohibition. The group would have only a system of shared empirical expectations on this matter, but that falls far short of a stable social morality. And, as I argued at the outset of this Part (§22.2), it is most unlikely that somehow this legal coercion will usher moral approval in its wake.

§29 Our Moral Nature and Governance in the Open Society

29.1. The “Theory of Self-Governance under Complexity?”

We have reviewed a great deal of our 3×3 space of governance. I have drawn on formal models, empirical evidence, and, I hope, some illuminating examples. Some may wish, after all these varied analyses, to know what is *the* theory of self-governance under complexity that I advocate. I hope, however, that it is reasonably clear that I do not propose any such theory—and I rather doubt that one is to be had. Governance is itself a diverse idea, and at different levels of complexity, different problems arise. This is not to say, however, that no conclusions have emerged. Let us briefly review some of our major conclusions.

Macro Governance

Of all the analyses, the strongest conclusions pertain to grand schemes whereby the governor guides society along a preferred path. There seems little prospect indeed of successful macro-level goal pursuit in a complex society. Disturbingly, this is precisely the sort of governance with which so much political philosophy, with its grand ambitions for general social transformation, has been obsessed. In this sense, our analysis constitutes a deep critique of the core project of a great deal of political philosophy. The ambitions of so much political philosophy—guiding society to a fully just, egalitarian, libertarian, socialist, or capitalist destination—are almost surely unachievable in a complex world. In this sense, complexity defeats some of the most enduring aims of political thinking.

At the macro level, I have argued, a governor, in cooperation with the self-organized normative framework of society, may well effectively shape the rules of self-organization. Helping to provide an institutional framework that “hinders hindrances” to self-organization—assisting to moderate destructive cascades while

avoiding “lock ins” on current practices—is a critical task of a governor in a complex society. But such institutional construction cannot take the form of “algorithms” that endeavor to provide rules that reliably secure targeted social outcomes. Given excellent rules of the game, it can be played in diverse and surprising ways. The institutional framework is not properly understood as “rules of regulation” that secure known ends: it provides a framework that allows an amazing array of social states, many of which we cannot yet imagine. We can seek to ensure that the rules are normatively acceptable to the public, but we cannot know to where they will lead us.

Meso Governance

Our conclusions are most ambiguous concerning meso, or mid-level, governance. Overall subsystem goal pursuit looks like a dubious enterprise. While the subsystem is somewhat simpler than a macro system, unless the system is significantly modular it is connected to other subsystems, and so many of the factors that determine the subsystem’s behavior emanate from outside, undermining efforts at subsystem control. Insofar as the Open Society involves autocatalytic diversity, producing ever-greater connected diversity, we would not expect this to be the norm.

Perhaps the sharpest disagreement concerns the effectiveness of a governor pursuing sectoral goals—environmental, economic, or welfare-targeted variables. We commenced with relatively simple many-to-one cases, where a target variable is affected by several others. Evidence indicates that the best forecasters can make accurate predictions in the short and medium term of such targets (up to about three years and in some cases five years), though in the medium to long term, prediction reduces to guessing. This, we have seen, is consistent with complexity. Some believe this cognitive accomplishment is sufficient for policy, but we must remember that there is a tremendous gap between prediction and control: problems of reflexivity and the lack of precise policy tools can undermine sectoral policy, even given sound predictions. Moreover, complex systems are nonlinear and are subject to change directions (think again of our simulation in §27.4); we have seen that in the medium to long term our target variables may “inexplicably” shift directions despite our short-term interventions. If sectoral policy has a desired trajectory of, say, ten years, it is dubious that it can be reliably achieved.

The deep problem, however, is that, though our policy thinking usually focuses on many-to-one relations, complex systems abound in many-to-one-to-many relations: one policy is an input into another, which ultimately may be an input into the first. Such feedbacks easily make the prediction of the effects of a policy incalculable. Add to this that our target variables are themselves usually systems of related variables, and we can begin to see why so much sectoral goal pursuit is a hit-and-miss affair. Despite all of this, we continue to treat our social world as essentially linear and determinate, selectively focusing on cases where the outcomes were welcomed, and putting down disappointments to failures of implementation, systematic injustice, or simply bad will.

Practically, meso-level goal pursuit in a complex world will usually be a matter of “muddling through” or, rather more impressively, learning-based governance. If so

much is beyond our ken, and the variables in theoretical models are (as Hayek so often insisted) unable to be adequately specified, perhaps the best the meso-level policymaker can do is to focus on a few variables, and seek to climb gradients, revaluing the moves at each point as she works toward locally better solutions. Whether or not such local incremental searching can achieve the desired results depends on the simplicity of the policy landscape, whether it is “dancing” with other policy landscapes, the signal-to-noise ratio (whether the policymaker has sufficiently good feedback to determine whether the previous move made things better or worse), and the degree of reflexivity. My own view is that these conditions are jointly demanding, and it is hard to see how they can be ignored. But we should draw back from sweeping claims of whether such policy is, as such, a good or a bad bet: we need to look at any given policy area and determine whether the conditions are adequately met.

Polycentric Governance

By its very nature, polycentric governance spans the various levels of policy, insisting that the right level of policy intervention depends on the public that shares a pressing strategic dilemma. I have suggested that as we “zoom out” to problems concerning very large publics, there will often be disagreement among citizens over whether the governor is helping to solve their strategic dilemma or pursuing its own goals. To the extent that this is the case, at these macro and higher meso levels, we can expect the policy to only imperfectly be addressing a shared pressing problem-solving context, in which citizens agree that they face an important problem, agree that they wish to solve it, are willing to pay the costs, and are willing to discount many unanticipated consequences in solving it. In such pressing problem-solving contexts, reflexivity, diversity, and complexity are greatly reduced, facilitating effective governance.

Polycentric governance studies have shown how institutional structuring is part and parcel of solving strategic dilemmas for publics that share pressing strategic dilemmas. The development of such institutional structures to facilitate meso- and micro-level joint action is by no means easy or certain, but is an important and often effective form of governance.

Micro Governance

One of the great shortcomings of political philosophy (as it has been of economics) is the paucity of attention paid to microlevel reform and policy. As I have noted, as political philosophers we are inclined to the “big questions,” and when we theorize about government and policy we naturally turn to the big questions of the justice of basic institutions, equity, freedom, and welfare. However, we now have a great deal of evidence that our attempts at grand policy are not terribly effective. As Banerjee and Duflo have taught us, patient and systematic studies evaluating attempts to improve our schools, hospitals, police forces, neighborhoods, unions, corporations, and universities are a much more promising route to truly effective governance than the grand projects of which we political philosophers are so enamored. Not only can these

interventions be systematically studied in random control trials, but they connect up much more easily with the underlying normative framework on which all social life depends. Let us consider this critical point more closely.

29.2. Governance and Our Moral Nature

We commenced our inquiry with Hayek's worry that our moral nature, which evolved in small groups, is a threat to the Open Society: we moralize about extensive complex systems as if they were, to use the unfortunate contemporary parlance, tribes. On the complicated picture that has emerged, we can see there is an insight here conjoined to a good bit of confusion and error. The insight is that human morality does focus on norm networks, and moral criticism and reform are often most effectively located in smaller networks. This has nothing to do with tribes, but has a great deal to do with shared mutual expectations and concern about what interacting others think one ought to do. This does not mean that we forgo worrying about big problems. As Banerjee and Duflo stressed, development economics has learned that to solve big problems, the governor usually must experiment and change smaller-scale ones. The two great errors of thinking about reform in complex societies is that it typically can be effectively top-down dictated, or that it is essentially impossible for the governor to do anything. By working with smaller-scale networks, important changes in the lives of citizens can be improved by the governor's participation in the self-organizing networks which are woven together in the Open Society. How all these changes affect each other seldom can be accurately predicted in the medium term, but improving schools, clinics, city services, and women's ability to effectively participate in economic life, while reducing debilitating disease from open defecation, domestic abuse, police brutality and bullying, official corruption, and sexual harassment are as unambiguously good goals as they come.

When changes come up from the more micro levels, not only are they apt to garner the moral endorsement of actual citizens, but the Open Society will possess a diversity of normative networks. Because what works today may be dysfunctional tomorrow, a diversity of approaches is always critical. This itself upsets the moralist, who believes she speaks for the truth about justice, and sees most deviations from her plan as shades of immorality. But many of the diverse publics will not take up her solutions—many citizens will see different problems and possibilities, and their normative beliefs will lead them to different solutions. We have seen that when this leads to deep polarization across many issues the moral order of the Open Society can be weakened, but within a large space, difference can strengthen it. In this way our moral nature, forming diverse moral networks, is the foundation of the Open Society. Rawls called a just society a “social union of social unions”¹—we might call the Open Society a network of moral networks.

¹ “The division of labor is overcome not by each becoming complete in himself, but by willing and meaningful work within a just social union of social unions in which all can freely participate as they so incline.” Rawls, *A Theory of Justice*, p. 464.

This largely completes our analysis of governance in complex systems. There is, however, one more issue to consider: the roles of expertise and democratic control in governance.

§30 Liberal Democracy

30.1. A Democracy of Experts

One of Hayek's recurring themes—it was the heart of his abuse of reason project—was that the moral sciences are not exercises in engineering.¹ For Hayek, the critical features of the engineering ideal are a command over the crucial variables, a planner who has sufficient knowledge of, and manipulative power over, these variables and their interactions to secure a given end. Of course, in actual engineering projects, these ideals are not fully met, but unless they can be reasonably approximated the project cannot go ahead. As Hayek repeatedly insisted, this engineering ethos was strongly attracted to economic planning. Vincent Ostrom convincingly argued that the picture of a top-down instrumentally rational pursuer of social goals was at the heart of the traditional theory of public administration emanating in the United States from Woodrow Wilson.² In this heavily expert-guided picture of society, democracy seems, as Schumpeter argued, reduced to periodic decisions about which elites and technocrats will run things.³ Perhaps the workers can select from a carefully edited menu about who will manage the nuclear power plant, but they certainly cannot have any real voice in running it.

If the socioeconomic system were simply complicated, rule by experts may be inevitable, as is civil engineering by experts. However, the complex is far beyond the complicated. As I have tried to demonstrate in this Part, it is not that there are no experts on complexity, but that the complexity they are expert on outruns their models, and their predictive and manipulative powers.⁴ As a response, theorists of complexity have looked to non-engineering models. To Hayek, the proper model is the gardener rather than the engineer—he thought that complexity theory allowed the “cultivation” of social processes.⁵ Colander and Kupers too argue that though policy cannot control society, it can “tweak” the evolutionary process.⁶ In many ways, the core point of this Part has been that the search for metaphors and general models is not useful: the policy space is large, and includes setting the rules of the game, solving strategic interactions, and pursuing goals at a variety of levels. There is no single task or role of the “policymaker.” Nevertheless, throughout most of this

¹ See his “Engineers and Planners.”

² Ostrom, *The Intellectual Crisis in American Public Administration*, chap. 2.

³ Schumpeter, *Capitalism, Socialism and Democracy*, chap. 22.

⁴ Colander and Kupers, *Complexity and the Art of Public Policy*, p. 175.

⁵ Hayek, “Degrees of Explanation,” pp. 209–10; “The Pretence of Knowledge,” pp. 371–2.

⁶ Colander and Kupers, *Complexity and the Art of Public Policy*, p. 59.

policy space, a recurring feature of complexity is that the dimensionality of the policy context is so high that no single expert model is apt to capture system dynamics. This reinforces Hayek’s insistence that no single expert has the knowledge to plan the system, since no single model of the system comes close to fully capturing its variables, their values, and their relations.

Indeed, Page has shown how a multiplicity of simple models may capture the crux of a complex system.⁷ As we have seen, perspectives include categorizations (§15.2); models based on different categorizations pick out different features of the world—or, more accurately, they categorize these features in different ways. Models based on different perspectives can simplify or “clump together” complex interaction effects; combining these simple models may approximate much more complex models. To better understand how, consider a slight modification of the toy example from Page.⁸ Suppose that we have a network comprising three variables (x, y, z), which simply have on/off values, as Display III.14. Suppose further that the values have an interaction effect, and the outcome is employment growth in a city.

x	y	z	Employment	Eq 6 prediction
0	0	0	0	.5
0	0	1	1	1
0	1	0	1	1
0	1	1	0	.5
1	0	0	2	1.5
1	0	1	1	1
1	1	0	1	1
1	1	1	2	1.5

Display III.14. Three-variable system with interaction effects (employment growth in 10K).

Even this simple interactive system is quite complicated to represent. The equation that maps these variables on to employment (E) is:

EQ. 3
$$E = 2x + y + z - 2(xy) - 2(xz) - 2(yz) + 4(xyz)$$

Now consider two simple linear models, based on perspectives A and B . Perspective A only has category x , and predicts on the basis of it alone:

EQ. 4
$$E_A = .5 + x.$$

⁷ Page, *The Difference*, pp. 214–30.

⁸ *Ibid.*, pp. 218–221.

In contrast, *B* seeks to model the interaction effects among all three categories (variables), but it simplifies this task by mapping strings of variables—*x*, *y*, *z* tuples—on to employment values, according to the rule:⁹

$$\text{EQ. 5} \quad E_A = \begin{cases} .5, & \text{if } x + y + z = 2 \\ 1.5, & \text{if } x + y + z = 0, 1, 3 \end{cases}$$

Both *A*'s and *B*'s models are simple; combining them, we get a model that, in fact, is surprisingly close to *E* (Eq. 3):

$$\text{EQ. 6} \quad E_{(\text{mean of } A + B)} = .5 + x + .5y + .5z - xy - xz - yz + 2xyz$$

As the fifth column in Display III.14 shows, the combined prediction of these simple models based on our two perspectives manages to catch most of the interaction effects. Of course they are not as good as a fully accurate model of all the interaction effects (Eq. 3), but the difficulty posed by complexity is precisely that such fully specified equations are often impossible to discover or calculate. Combining simpler models that handle the interactions in different ways (*A* ignores them, while *B* uses a simple function) can help us cope with complexity, but for that we need perspectives with very different categorizations and models.

Page dubs this “The Crowd’s Possibly Free Lunch Theorem.”¹⁰ But free lunches on complicated matters are not common: we should structure our searches so that we can get them when on offer, but we shouldn’t count on them. Wolpert and McCready’s “No Free Lunch Theorem” showed that when searching for solutions to difficult problems, not only is there no search strategy that is best for all problems, *but no search strategy is best on average*. A strategy that solves some problems well performs poorly on others. The upshot is that even if we restrict ourselves to a population of complexity experts, they must constantly draw on the “wisdom of expert crowds.” In complex problems, relying on the best expert, or the “best practice,” is a mistake. Diversity of models and procedures is required to understand complex systems and respond to the challenges they generate. If all adapt the best theory or the best approach, we homogenize our ability to respond, and like a beautiful, groomed monoculture, it can be wiped out by the first radical shift it confronts.

30.2. A Democracy of All

So the type of expertise characteristic of complexity science is seldom of the type to warrant an expert in seeking to direct policy. A community of experts is required. Now the revolutionary work of Lu Hong and Scott Page justifies a more radical conclusion: in a variety of problem-solving contexts, an increase of diversity more than

⁹ As Page (ibid., p. 391) points out, this is a Walsh function. See Page and Richardson, “Walsh Functions, Schema Variance, and Deception.”

¹⁰ *The Difference*, p. 221.

compensates for a decrease in expertise. Diversity, in short, trumps ability.¹¹ Experts tend to think alike, and so categorize problems in similar ways. To increase the ability to model (i.e., think about) more dimensions, we often need to bring in non-expert perspectives: diversity can lead to solutions where expertise is stumped. Some have been enthusiastic about the application of Hong and Page's work to democratic decision-making, while others have questioned its relevance.¹² My concern here is democracy in the widest sense: whether the nature of expertise in complex systems provides a case for rule by experts. It clearly does not. Importantly, the Hong-Page theorem has application to situations in which people agree about the contours of an adequate solution of a complex policy problem, but disagree about the best policy to secure it.¹³ As I have argued (§25), this is precisely the pattern of agreement and disagreement most characteristic of strategic dilemmas, which provide the focus of polycentric governance. Given this, problems of complexity point to a highly diverse, broadly democratic, approach to policy. It is almost diametrically opposed to the rule by elite experts.

More fundamentally, however, policymaking must be broadly democratic, as without the public justification secured through popular participation, policy is very likely to be ineffective and perhaps counterproductive. One reason why Hayek insisted that policymaking was entirely different than technical engineering was that the policymaker deals with reflexive agents who make choices of their own.¹⁴ The engineer does not have to worry that the steel may decide to violate the design specifications; the policymaker must always confront this possibility. Public administration has too often seen populations as passive materials, whose actions could be dictated by policy, rather than as self-governing agents.¹⁵ This passive population model has, amazingly, often supposed that people would act against their own judgments when directed or cajoled to do so. Unfortunately, this view has been resurrected by those elites who continue to believe that the public is too ignorant to make its own decisions, and so should submit to "epistocracy," or rule by those who know (aka, them).¹⁶ Not only, however, is such expertise essentially nonexistent in complex systems, but most actual agents in the Open Society are anything but passive materials to be guided by the elite: they are active, reflexive agents who make their own choices. When citizens do not endorse a policy, many will employ their resources to evade it. A state can criminalize recreational drugs, but it cannot stop illegal markets, drug cartels, and street violence from ruining nations; it can prohibit abortions, but cannot stop

¹¹ See Hong and Page, "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers"; Hong and Page, "Problem Solving by Heterogeneous Agents"; Page, *The Difference*, chap. 6.

¹² For enthusiasm, see Landemore, *Democratic Reason*. For skepticism about the application to electoral systems, see Weymark, "Cognitive Diversity, Binary Decisions, and Epistemic Democracy."

¹³ I have argued that it does not get much grip on political questions in which we deeply disagree about the goal of the policy, or what would constitute a successful policy. See *The Tyranny of the Ideal*, chap. 3.

¹⁴ Hayek, "Engineers and Planners," p. 158.

¹⁵ See Ostrom, *The Intellectual Crisis in American Public Administration*; Aligica, Boettke, and Tarko, *Public Governance and the Classical Liberal Perspective*.

¹⁶ For an excellent response, see Aligica, *Public Entrepreneurship and Self-Governance*, Part II. Often these experts claim authority on the basis of highly dubious economic models. The history of the influence of economic experts has been rather unsettling. See Levy and Peart, *Escape from Democracy: The Role of Experts and the Public in Economic Policy*.

women from doing their best to secure them; it can supertax the rich, but cannot prevent them from moving their resources to other jurisdictions—or prevent those with lower incomes from supporting the rich “against their own interests”; it can use force against undocumented immigrants, but cannot stop them from crossing the border, or employers from hiring them; it can prohibit corruption, female genital cutting, and child marriage, but when the underlying norms support them, it can do little; it can censor the press, but cannot prevent people from finding alternative sources of information. In all the cases it can thwart some, but many find ways around policies they loathe, and the state itself loses legitimacy—unless, that is, like contemporary China, it is prepared to use extreme and sustained force to impose its “epistocratic” goals to intimidate large parts of the population.

30.3. Freedom and Democracy

Recall Mussolini’s claim: “the more complicated the forms assumed by civilization, the more restricted the freedom of the individual must become.”¹⁷ In one way he was right: if society were merely complicated, then as with any complicated mechanism, the functioning of the parts must be constrained by the requirements of the whole. Having parts of a nuclear power plant deciding how to go about their business is not a good design principle. Fascism or communism is an appropriate ideology for society as a machine. The expert engineers must be given the power to control, repair, and redesign the complicated machinery of society, while the cogs, pullies, engines, and valves—aka the citizens—must perform according to their assigned roles.

But, again, the complex cannot be understood in terms of the complicated. The Open Society is, at bottom, a self-organized order. Agents must have the freedom to reflexively adjust their actions to the decisions of others, as people explore the adjacent possible, and discover new niches (ways of life, services, and products) that yesterday could not be even conceived. In an Open Society, individuals come together at various levels to solve collective problems, but always in the context of a system in which one group’s democratic decisions require reflexive adjustments by others. Individual freedom is not a limit on democracy, it is a necessary part of the process of adjustment and change that includes democracy itself. As liberal democrats have long insisted, so far from being opposed or in tension, democracy and freedom need each other to thrive.

Without democracy, individual reflexive action leads to strategic dilemmas, externalities, and unfair structures of interaction. We must remember that the human urge to dominate remains. The Open Society has evolved an increasingly impartial and wide-ranging system of rules and norms. A critical task of the democratic order is to ensure the equality and fairness on which large-scale human cooperation depends. Without these rules of the game, self-organization can lead to oppressive hierarchies.

On the other hand, without extensive freedom to explore and adjust, democracy leads to group conflict, rigidity, and a reverse dominance hierarchy. Unless it is

¹⁷ Hayek, *The Road to Serfdom*, p. 91. See note 7 of the Prolegomenon.

animated by a spirit of public justification, democracy itself becomes a mechanism by which some seek to impose their valued goals on others in the name of the people. The democracy at the heart of the Open Society is not that which is popular among contemporary political philosophers, in which the state articulates values that “we” share, and which are valorized into “our” goals. When this occurs, the state, in the eyes of many, simply becomes a coercive agent pursuing exogenous goals, which they must work around or protest against. A healthy Open Society can certainly tolerate such attempts to impose controversial values in the name of all, but we should not expect these efforts to yield the hoped-for ends. In any event, that is not the sort of democracy that the Open Society requires, even if it can be coped with. Effective governance requires widely justified norms and policies. Here, Hayek was deeply mistaken: without justification, the Open Society could not self-organize.

Epilogue

I shall not attempt to catalogue my responses to Hayek's three unsettling theses. My purpose has not been to provide a checklist of where he was correct and where, at least in my view, we should depart from his analyses. What would be the point in that? Throughout, my aim has been to follow out Hayek's research program, learning from him, but always being alert to where we should revise this revolutionary agenda for social philosophy. The result, admittedly, has been rather complex (in a more mundane sense), but simple answers and unambiguous conclusions are highly suspect when thinking about whether we are fit for the Open Society, whether we can justify it, and whether we can govern it. If one presses for an overall evaluation, I would say that Hayek's program was manifestly prescient and strikingly sophisticated, always identifying real and pressing problems, though he often underestimated the resources of human morality and the Open Society to cope with the challenges he perceived. Unlike so much contemporary political philosophy, we can learn from, and improve upon, his remarkable work. I no doubt have made (more than) my own share of errors in this work—my hope is that they help spur the next generation of social philosophers to do better.

A core theme of this work has been the moral foundations of human cooperation, from the Late Pleistocene era to the Open Society. Far too often, social theory has assumed that self-interest is a building block of cooperation: morality is all well and good, but cooperation, especially among diverse individuals, must be based on some equilibrium of interests. Morality requires the cozy agreement of the tribe, while interests scale up to organize a diverse world. Parts I and II sought to demonstrate that this apparently hard-headed, realistic attitude has strikingly little support. Diversity, far from being the enemy of a society with a shared morality, can be an engine for it. Self-organization and public justification can be powered by diversity as the morality of the Open Society is pushed up the ladder of impartiality. We are neither unfit for this Open Society nor optimized for it. The ancient human abilities to cooperate through norms and learn from others are the foundations of its autocatalytic order, and our egalitarianism is closely bound up with our readiness to adopt impartial rules and reject domineering ones. Yet we have always been imperfect and ambivalent cooperators: extending moralized impartial cooperation has always tussled with the forces of egoism, domination, and the attractions of deeper relations with kin and ethnic networks.

Hayek was correct that we have created a society that cannot be navigated to arrive at a destination. Throughout his career, he criticized grand plans for macro guidance, from socialist planning to theories of social justice that specified distributional outcomes. He has been harshly criticized for his critique, but everything we have considered in Part III indicates that it was essentially sound. However, that governance cannot be a grand project in social planning does not imply that a complex society cannot be governed. At every level—the macro, meso, and micro—we have seen that

governors, in conjunction with non-state actors and norm networks, can reform the system, help solve pressing problems, and work to improve people's lives. While a political theory that appreciates complexity cannot chart courses for us, it can help us see when governance schemes for a complex society are likely to be a will-o'-wisp, and when they can be effective in improving people's lives. In conjunction with social forces, democratic governance can render the Open Society more resilient, help solve myriad strategic dilemmas, and improve the lives of its citizens in ways that are as important as they appear modest. Yet much remains outside the ambit of governance and must be left to the self-organizing forces that drive the Open Society. To ignore these limits can only bring democratic governance into disrepute: promising a new social order, it delivers only disappointment, cynicism, and populism.

As Hayek, Popper, and Schumpeter taught us, there are numerous forces that resist the Open Society. Hayek repeatedly stressed how, enamored with a simple nineteenth-century view of science, elites believe that with enough power they can guide it. Popper was certainly correct that those who dream of ideals and utopias are usually its enemies—and this, alas, includes far too many philosophers. And Schumpeter was correct that it excels at subsidizing its own critics.¹ Many are suspicious of it for upsetting the present and familiar, which they cherish, as it constantly expands the world of cooperation and impartial, inclusive morality. Others believe that protecting their co-nationals requires closure.² And, most understandably, many are unsettled by its waves of creative and uncreative destruction. Indeed, it is probably part of the expected tempo of human society to move toward openness and impartiality when more secure, and draw back toward closed parochialism in times of storm and stress. As Buchanan and Powell observe, “Conditions of infectious disease, physical insecurity, interethnic conflict, and low rates of productivity seed exclusivist moral responses, which in turn feed back into the exacerbation and perpetuation of the conditions that trigger exclusivist tendencies.”³ The years following 2020 are apt to manifest such a contraction phase, perhaps a severe one.

However, given the autocatalytic nature of the Open Society, it is an ever-increasing and relentless engine of diversity and inclusivity. It can be slowed and can even wane for a time, but only great and sustained coercion can truly put the brake on it. When some avenues are blocked, the Open Society's niche creation and innovation manifest themselves elsewhere. Even those who are willing to employ great and systematic coercion to control it find that as they block one avenue of diversity, others crop up. Unlike in the story of the Little Dutch Boy who stopped a flood by putting his finger in a leaking dike, the enemy of the Open Society cannot stop it by halting this or that exploration.⁴ For a time, opponents of diversity and openness may check exploration,

¹ Schumpeter, *Capitalism, Socialism and Democracy*, chaps. VIII, XI.

² It is often claimed, by both the left and the right, that a restrictive immigration policy is required to defend the interests of American workers. The data, however, do not indicate that immigration substantially harms American workers overall; the effects on certain subpopulations is more controversial, but even in these cases (such as high school dropouts) some studies find almost no effect. National Academies of Sciences, Engineering, and Medicine, *The Economic and Fiscal Consequences of Immigration*, p. 267.

³ Buchanan and Powell, *The Evolution of Moral Progress*, p. 210.

⁴ “You see, when a dike is about to break, a finger just does not cut it. Dikes don't typically leak—they weaken until whole sections are washed away. No finger will help when that happens. So what is a [populist] hero supposed to do? You don't use a finger, you use a boat.” Hoitink, “Why the Little Dutch Boy Never Put His Finger in the Dike.”

creativity, diversity, and freedom, but the forces propelling the new civilization proclaimed by Popper and Hayek cannot long be contained. The real question is whether we can understand the possibilities and constraints of this complex civilization, or struggle against it, seeking a mastery that will elude us.

APPENDIX A

Agent Types

<i>N</i>	Moderately Conditional Cooperators Weighting	Linear Agents Weighting	Quasi-Kantians Weighting	Highly Conditional Cooperators Weighting
1	0	0	0	0
2	0	0.01	0	0
3	0	0.03	0	0
4	0	0.03	0	0
5	0	0.04	0	0
6	0	0.05	0	0
7	0	0.06	0	0
8	0	0.07	0	0
9	0	0.08	0	0
10	0	0.09	0.02	0
11	0	0.1	0.04	0
12	0	0.11	0.06	0
13	0	0.12	0.08	0
14	0	0.13	0.11	0
15	0	0.14	0.13	0
16	0	0.15	0.15	0
17	0	0.16	0.17	0
18	0	0.17	0.2	0
19	0	0.18	0.24	0
20	0	0.19	0.29	0
21	0	0.2	0.34	0
22	0	0.21	0.39	0
23	0	0.22	0.42	0
24	0	0.23	0.45	0
25	0	0.24	0.47	0
26	0	0.25	0.51	0
27	0	0.26	0.53	0
28	0	0.27	0.55	0
29	0	0.28	0.58	0
30	0	0.29	0.61	0
31	0.01	0.3	0.62	0
32	0.06	0.31	0.63	0
33	0.11	0.32	0.65	0
34	0.15	0.33	0.67	0
35	0.19	0.34	0.69	0.01
36	0.21	0.35	0.71	0.015
37	0.23	0.36	0.72	0.017
38	0.24	0.37	0.73	0.0175

<i>N</i>	Moderately Conditional Cooperators Weighting	Linear Agents Weighting	Quasi-Kantians Weighting	Highly Conditional Cooperators Weighting
39	0.27	0.38	0.75	0.018
40	0.31	0.39	0.77	0.0185
41	0.33	0.4	0.79	0.019
42	0.35	0.41	0.81	0.0195
43	0.38	0.42	0.83	0.02
44	0.41	0.43	0.85	0.03
45	0.45	0.44	0.87	0.035
46	0.48	0.45	0.89	0.04
47	0.51	0.46	0.91	0.045
48	0.54	0.47	0.94	0.05
49	0.57	0.48	0.97	0.055
50	0.6	0.49	1	0.06
51	0.62	0.5	1	0.065
52	0.64	0.51	1	0.07
53	0.66	0.52	1	0.075
54	0.68	0.53	1	0.085
55	0.7	0.54	1	0.095
56	0.72	0.55	1	0.09
57	0.74	0.56	1	0.095
58	0.76	0.57	1	0.1
59	0.78	0.58	1	0.11
60	0.8	0.59	1	0.12
61	0.81	0.6	1	0.13
62	0.82	0.61	1	0.14
63	0.83	0.62	1	0.15
64	0.84	0.63	1	0.16
65	0.85	0.64	1	0.17
66	0.86	0.65	1	0.18
67	0.87	0.66	1	0.19
68	0.88	0.67	1	0.2
69	0.89	0.68	1	0.21
70	0.9	0.69	1	0.23
71	0.91	0.7	1	0.25
72	0.92	0.71	1	0.27
73	0.93	0.72	1	0.29
74	0.94	0.73	1	0.31
75	0.95	0.74	1	0.33
76	0.96	0.75	1	0.35
77	0.97	0.76	1	0.37
78	0.98	0.77	1	0.39
79	0.99	0.78	1	0.41
80	1	0.79	1	0.43
81	1	0.8	1	0.45
82	1	0.81	1	0.47
83	1	0.82	1	0.49
84	1	0.83	1	0.51
85	1	0.84	1	0.53

<i>N</i>	Moderately Conditional Cooperators Weighting	Linear Agents Weighting	Quasi-Kantians Weighting	Highly Conditional Cooperators Weighting
86	1	0.85	1	0.56
87	1	0.86	1	0.59
88	1	0.87	1	0.62
89	1	0.88	1	0.63
90	1	0.89	1	0.66
91	1	0.9	1	0.69
92	1	0.91	1	0.72
93	1	0.92	1	0.75
94	1	0.93	1	0.78
95	1	0.94	1	0.81
96	1	0.95	1	0.85
97	1	0.96	1	0.88
98	1	0.97	1	0.91
99	1	0.98	1	0.94
100	1	0.99	1	0.97
101	1	1	1	1

APPENDIX B

Equations in Simulation §27

Given the set of variables, $\{a, b, c, d, e, f, g, h\}$, where the first four are targeted with 1 unit of investment per period, and where x_{i-1} is the value of any variable in the previous period, the equations for each variable's value in period i are (starting with an initial value in period 1 of 5 units):

Targeted Variables

$$a_i = a_{i-1} + 1 - (.1 \times e_{i-1}) - (.2 \times b_{i-1})$$

$$b_i = b_{i-1} + 1 - (.2 \times f_{i-1}) - (.1 \times e_{i-1})$$

$$c_i = c_{i-1} + 1 + (.1 \times a_{i-1}) + (.2 \times b_{i-1})$$

$$d_i = d_{i-1} + 1 + (.1 \times a_{i-1}) - (.2 \times b_{i-1})$$

Non-targeted Variables

$$e_i = x_{i-1} + (.1 \times c_{i-1}) + (.1 \times g_{i-1})$$

$$f_i = f_{i-1} - (.1 \times c_{i-1}) + (.2 \times g_{i-1})$$

$$g_i = g_{i-1} - (.2 \times d_{i-1}) + (.2 \times h_{i-1})$$

$$h_i = h_{i-1} - (.1 \times d_{i-1}) - (.2 \times g_{i-1})$$

Bibliography

- Abraham, Anita, and Jean-Philippe Platteau. "Participatory Development in the Presence of Endogenous Community Imperfections." Department of Economics and CRED (Centre de Recherche en Economie du Développement), University of Namur, 2002.
- Alexander, J. McKenzie. *The Structural Evolution of Morality*. Cambridge: Cambridge University Press, 2007.
- Alexander, Richard D. *Human Social Evolution*, edited by Kyle Summers and Bernard Crespi. Oxford: Oxford University Press, 2013.
- Aligica, Paul Dragos. *Institutional Diversity and Political Economy*. Oxford: Oxford University Press, 2014.
- Aligica, Paul Dragos. *Public Entrepreneurship, Citizenship and Self-Governance*. Cambridge: Cambridge University Press, 2018.
- Aligica, Paul Dragos, and Peter J. Boettke. *Challenging Institutional Analysis and Development*. New York: Routledge, 2009.
- Aligica, Paul Dragos, Peter J. Boettke, and Vlad Tarko. *Public Governance and the Classical Liberal Perspective: Political Economy Foundations*. Oxford: Oxford University Press, 2019.
- Anderson, Elizabeth S. "John Stuart Mill and Experiments in Living." *Ethics*, vol. 102 (1991): 4–26.
- Anderson, Elizabeth S. "Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain's Abolition of Slavery." The Lindley Lecture, University of Kansas, February 11, 2014. Available at: https://kuscholarworks.ku.edu/bitstream/handle/1808/14787/Anderson_Social_Movements.pdf.
- Andrighetto, Giulia, Daniel Villatoro, and Rosaria Conte. "Norm Internalization in Artificial Societies." *AI Communications*, vol. 23 (2010): 325–39.
- Appiah, Kwame Anthony. *As If: Idealizations and Ideals*. Cambridge, MA: Harvard University Press, 2017.
- Appiah, Kwame Anthony. *Cosmopolitanism: Ethics in a World of Strangers*. New York: W. W. Norton, 2006.
- Arthur, W. Brian. *Complexity and the Economy*. Oxford: Oxford University Press, 2015.
- Arthur, W. Brian. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press, 1994.
- Auyang, Sunny Y. *Foundations of Complex-Systems Theories in Economics, Evolutionary Biology and Statistical Physics*. Cambridge: Cambridge University Press, 1998.
- Axelrod, Robert. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press, 1997.
- Axelrod, Robert. "The Emergence of Cooperation Among Egoists." *The American Political Science Review*, vol. 75 (1981): 306–18.
- Axelrod, Robert. *The Evolution of Cooperation*. New York: Basic Books, 1984.
- Axelrod, Robert, and Michael D. Cohen. *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. New York: Basic Books, 2000.

- Axtell, Robert, et al. "Challenges of Integrating Complexity and Evolution in Economics." In *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, 2016: 65–84.
- Ayars, Alisabeth, and Shaun Nichols. "Rational Learners and Metaethics: Universalism, Relativism, and Evidence from Consensus." *Mind and Language*, vol. 35 (2019): 1–23. doi: 10.1111/mila.12232.
- Baier, Kurt. *The Moral Point of View: A Rational Basis of Ethics*. Ithaca, NY: Cornell University Press, 1958.
- Baier, Kurt. *The Moral Point of View: A Rational Basis of Ethics*, abridged ed. New York: Random House, 1965.
- Baier, Kurt. *The Rational and the Moral Order*. Chicago: Open Court, 1995.
- Banerjee, Abhijit V., and Esther Duflo. "The Economic Lives of the Poor." *Journal of Economic Perspectives*, vol. 21 (Winter 2007): 141–67.
- Banerjee, Abhijit V., and Esther Duflo. *Good Economics for Hard Times*. New York: Public Affairs, 2019.
- Banerjee, Abhijit V., and Esther Duflo. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs, 2011.
- Barrett, Jacob, and Gerald Gaus. "Laws, Norms, and Public Justification: The Limits of Law as an Instrument of Reform." In *Public Reason and Courts*, edited by Silje A. Langvatn, Mattias Kumm, and Wojciech Sadurski. Cambridge: Cambridge University Press, 2020: 201–28.
- Barry, Brian. "John Rawls and the Search for Stability." *Ethics*, vol. 105 (1995): 874–915.
- Barry, Brian. *Justice as Impartiality*. Oxford: Oxford University Press, 1995.
- Beinhocker, Eric D. *The Origin of Wealth: Evolution, Complexity and the Radical Remaking of Economics*. Cambridge, MA: Harvard Business School Press, 2006.
- Beinhocker, Eric D. "Reflexivity, Complexity, and the Nature of Social Science." *Journal of Economic Methodology*, vol. 20 (2013): 330–42.
- Belot, Michèle, Vincent P. Crawford, and Cecilia Heyes. "Players of Matching Pennies Automatically Imitate Opponents' Gestures against Strong Incentives." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110 (February 19, 2013): 2763–8.
- Benn, S. I. *A Theory of Freedom*. Cambridge: Cambridge University Press, 1988.
- Benn, S. I., and G. F. Gaus, eds. *Public and Private in Social Life*. New York: St. Martin's Press, 1983.
- Bhattacharyya, M. N., and Allan P. Layton. "Effectiveness of Seat Belt Legislation on the Queensland Road Toll." *Journal of the American Statistical Association*, vol. 74 (September 1974): 596–603.
- Bicchieri, Cristina. *The Grammar of Society*. Cambridge: Cambridge University Press, 2005.
- Bicchieri, Cristina. *Norms in the Wild*. Oxford: Oxford University Press, 2016.
- Bicchieri, Cristina, and Alex Chavez. "Norm Manipulation, Norm Evasion: Experimental Evidence." *Economics and Philosophy*, vol. 29 (July 2013): 175–98.
- Bicchieri, Cristina, Eugen Dimant, Simon Gächter, and Daniele Nosenzo. "Social Proximity and the Evolution of Norm Compliance." *SSRN Electronic Journal* (January 2019). doi: 10.2139/ssrn.3355028.

- Bicchieri, Cristina, Eugen Dimant, and Erte Xiao. "Deviant or Wrong? The Effects of Norm Information on the Efficacy of Punishment." monash.edu/business-economics. ISSN number 1441-5429. Discussion number 07/18.
- Bicchieri, Cristina, and Hugo Mercier. "Norms and Beliefs: How Change Occurs." *Iyyun: The Jerusalem Philosophical Quarterly*, vol. 63 (January 2014): 60–82.
- Bicchieri, Cristina, and Peter McNally. "Shrieking Sirens—Schemata, Scripts, and Social Norms: How Change Occurs." *Social Philosophy & Policy*, vol. 35 (Spring 2018): 23–53.
- Bicchieri, Cristina, and Hugo Mercier. "Self-Serving Biases and Public Justifications in Trust Games." *Synthese*, vol. 190 (2013): 909–22.
- Bicchieri, Cristina, and Erte Xiao. "Do the Right Thing: But Only If Others Do So." *Journal of Behavioral Decision Making*, vol. 22 (2009): 191–208.
- Bierhoff, Hans-Werner, and Elke Rohmann. "Conditions for Establishing a System of Fairness: Comment on Brosnan." *Social Justice Research*, vol. 19 (June 2006): 194–200.
- Bilgrami, Akeel. "Gandhi, the Philosopher." Available at: <https://philosophy.columbia.edu/files/philosophy/content/BilgramiGandhi.pdf>.
- Binmore, Ken. *Crooked Thinking or Straight Talk? Modernizing Epicurean Scientific Philosophy*. Gewerbestrasse, Switzerland: Springer, 2020.
- Binmore, Ken. *Game Theory and the Social Contract: Just Playing*. Cambridge, MA: MIT Press, 1998.
- Binmore, Ken. *Game Theory and the Social Contract: Playing Fair*. Cambridge, MA: MIT Press, 1995.
- Binmore, Ken. *Natural Justice*. Oxford: Oxford University Press, 2005.
- Binmore, Ken, and Avner Shaked. "Experimental Economics: Where Next?" *Journal of Economic Behavior and Organization*, vol. 73 (2010): 87–100.
- Boehm, Christopher. "Ancestral Hierarchy and Conflict." *Science*, vol. 336 (May 2012): 844–7.
- Boehm, Christopher. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press, 1999.
- Boehm, Christopher. *Moral Origins: The Evolution of Virtue, Altruism and Shame*. New York: Basic Books, 2012.
- Boettke, Peter J. F. A. *Hayek: Economics, Political Economy and Social Philosophy*. New York: Palgrave Macmillan, 2018.
- Böhm, Robert, Isabel Thielmann, and Benjamin E. Hilbig. "The Brighter the Light, the Deeper the Shadow: Morality Also Fuels Aggression, Conflict, and Violence." *Behavioral and Brain Sciences*, vol. 41 (2018): 15–16.
- Bosanquet, Bernard. "The Antithesis between Individualism and Socialism Philosophically Considered." In *The Philosophical Theory of the State and Related Essays*, edited by Gerald Gaus and William Sweet. South Bend, IN: St. Augustine's Press, 2001: 224–46.
- Bosanquet, Bernard. "The Philosophical Theory of the State." In *The Philosophical Theory of the State and other Essays*, edited by Gerald Gaus and William Sweet. South Bend, IN: St. Augustine's Press, 2001: 1–293.
- Bowles, Samuel, and Herbert Gintis. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton University Press, 2011.
- Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. "The Evolution of Altruistic Punishment." In *Moral Sentiments and Material Interests: The Foundations of*

- Cooperation in Economic Life*, edited by Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr. Cambridge, MA: MIT Press, 2005: 215–28.
- Boyd, Robert, and Peter J. Richerson. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press, 1985.
- Boyd, Robert, and Peter J. Richerson. “Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups.” In *The Origin and Evolution of Cultures*, edited by Boyd and Richerson. New York: Oxford University Press, 2005: 166–88.
- Boyd, Robert, and Peter J. Richerson. “Shared Norms and the Evolution of Ethnic Markers.” In *The Origin and Evolution of Cultures*, edited by Boyd and Richerson. New York: Oxford University Press, 2005: 118–31.
- Boyd, Robert, and Peter J. Richerson. “Simple Models of Complex Phenomena.” In *The Origin and Evolution of Cultures*, edited by Boyd and Richerson. New York: Oxford University Press, 2005: 397–419.
- Boyd, Robert, and Peter J. Richerson. “Why Does Human Culture Increase Human Adaptability?” In *The Origin and Evolution of Cultures*, edited by Boyd and Richerson. New York: Oxford University Press, 2005: 35–51.
- Brandt, Richard B. *Hopi Ethics: A Theoretical Analysis*. Chicago: University of Chicago Press, 1954.
- Brosnan, Sarah F. “Nonhuman Species’ Reactions to Inequity and Their Implications for Fairness.” *Social Justice Research*, vol. 19 (June 2006): 153–85.
- Buchanan, Allen, and Russell Powell. *The Evolution of Moral Progress: A Biocultural Theory*. New York: Oxford University Press, 2018.
- Buchanan, James. *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press, 1975.
- Buchanan, James. “Taxation in Fiscal Exchange.” In *The Collected Works of James M. Buchanan*, vol. I: *The Logical Foundations of Constitutional Liberty*, with a Forward by Geoffrey Brennan, Hartmut Kliemt, and Robert D. Tollison. Indianapolis, IN: Liberty Fund, 1999: 133–49.
- Buchanan, James, and Gordon Tullock. *The Calculus of Consent*. Ann Arbor: University of Michigan Press, 1962.
- Buharin, N., and E. Preobrazhensky. *The ABC of Communism: A Popular Explanation of the Program of the Communist Party of Russia*, translated by Eden and Cedar Paul. The Communist Party of Great Britain, 1922.
- Burke, Edmund. *Select Writings of Edmund Burke*. Indianapolis, IN: Liberty Fund, 1999.
- Byrne, Richard W., and Andrew Whiten. “Machiavellian Intelligence.” In *Machiavellian Intelligence II: Extensions and Evaluations*, 2nd ed., edited by Richard W. Byrne and Andrew Whiten. Cambridge: Cambridge University Press, 1997: 1–23.
- Cairney, Paul. “Complexity Theory in Political Science and Public Policy.” *Political Studies Review*, vol. 10 (2012): 346–58.
- Caldwell, Bruce. *Hayek’s Challenge: An Intellectual Biography of F. A. Hayek*. Chicago: University of Chicago Press, 2004.
- Call, Josep, Malinda Carpenter, and Michael Tomasello. “Copying Results and Copying Actions in the Process of Social Learning: Chimpanzees (*Pan troglodytes*) and Human Children (*Homo sapiens*).” *Animal Cognition*, vol. 8 (2005): 151–63.
- Caplan, Bryan. “Two Takes on Hayek’s Flabbiness.” *Econlog*, Dec. 13, 2009. https://www.econlib.org/archives/2009/12/two_takes_on_ha.html, accessed August 23, 2018.

- Carnes, Joseph H. *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*. Chicago: University of Chicago Press, 1981.
- Cartwright, Nancy, and Jeremy Hardie. *Evidence-Based Policy*. Oxford: Oxford University Press, 2012.
- Chapman, John W. "Toward a General Theory of Human Nature and Dynamics." In *NOMOS XVII: Human Nature in Politics*, edited by J. Roland Pennock and John W. Chapman. New York: New York University Press, 1977: 292–319.
- Chaudhuri, Ananish. "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature." *Experimental Economics*, vol. 14 (2011): 47–83.
- Chwe, Michael Suk-Young. *Rational Ritual: Culture, Coordination and Common Knowledge*. Princeton, NJ: Princeton University Press, 2001.
- Chung, Hun. "Hobbes' State of Nature: A Modern Bayesian Game-Theoretic Analysis." *Journal of the American Philosophical Association*, vol. 1 (2015): 485–508.
- Cialdini, Robert B., Carl A. Kallgren, and Raymond B. Reno. "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior." *Advances in Experimental Social Psychology*, vol. 24 (1990): 201–34.
- Cislaghi, Beniamino, Diane Gillespie, and Gerry Mackie. *Values Deliberation and Collective Action in Rural Senegal*. Wallace Global Fund and UNICEF Child Protection Section, 2014.
- Coase, R. M. "The Nature of the Firm." In his *The Firm, the Market and the Law*. Chicago: University of Chicago Press, 1998: chap. 2.
- Coase, R. H. "The Problem of Social Cost." *Journal of Law and Economics*, vol. 3 (1960): 1–44.
- Cochran, Gregory, and Henry Harpending. *The 10,000 Year Explosion: How Civilization Accelerated Human Evolution*. New York: Basic Books, 2009.
- Cohen, Alma, and Liran Einav. "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities." *Review of Economics and Statistics*, vol. 85 (2003): 828–43.
- Cohen, G. A. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press, 2008.
- Cohen, G. A. *Why Not Socialism?* Princeton, NJ: Princeton University Press, 2009.
- Colander, David, and Roland Kupers. *Complexity and the Art of Public Policy: Solving Society's Problems from the Bottom Up*. Princeton, NJ: Princeton University Press, 2014.
- Cook, Richard, Geoffrey Bird, Gabriele Lünser, Steffen Huck, and Cecilia Heyes. "Automatic Imitation in a Strategic Context: Players of Rock—Paper—Scissors Imitate Opponents' Gestures." *Proceedings of the Royal Society B: Biological Sciences*, vol. 279, no. 1729 (February 22, 2012): 780–6.
- Cosmides, Leda, and John Tooby. "Cognitive Adaptations for Social Exchange." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by Jerome H. Barkow, Leda Cosmides, and John Tooby. New York: Oxford University Press, 1992: 163–228.
- Crowder, George. "Pluralism." In *The Routledge Companion to Social and Political Philosophy*, edited by Gerald Gaus and Fred D'Agostino. New York: Routledge, 2013: 353–63.
- Cummins, Denise Dellaros. "Evidence for the Innateness of Deontic Reasoning." *Mind & Language*, vol. 11 (June 1996): 160–90.

- Cummins, Denise Dellarosa. "Evidence of Deontic Reasoning in 3- and 4-Year-Olds." *Memory and Cognition*, vol. 24 (1996): 823–9.
- D'Agostino, Fred. "From the Organization to the Division of Cognitive Labor." *Politics, Philosophy & Economics*, vol. 8 (2009): 101–29.
- D'Agostino, Fred. "How Can We Do Political Philosophy?" *Cosmos + Taxis* (January 2018): 29–37.
- D'Agostino, Fred. *Naturalizing Epistemology*. New York: Palgrave, 2010.
- D'Agostino, Fred. "The Orders of Public Reason." *Analytic Philosophy*, vol. 54 (2013): 129–55.
- Darwin, Charles. *The Descent of Man and Selection in Relation to Sex*. In *The Origin of Species and the Descent of Man*. New York: Modern Library, n.d.: 387–934.
- Darwin, Charles. *The Origins of Species*. In *The Origin of Species and the Descent of Man*. New York: Modern Library, n.d.: 3–386.
- De Vlieghere, Martin. "A Reappraisal of Friedrich A. Hayek's Cultural Evolutionism." *Economics and Philosophy*, vol. 10 (1994): 285–304.
- de Waal, Frans. *The Bonobo and the Atheist: In Search of Humanism among the Primates*. New York: Norton, 2013.
- de Waal, Frans. *Chimpanzee Politics: Power and Sex among Apes*, 25th anniversary ed. Baltimore, MD: Johns Hopkins University Press, 2007.
- DellaVigna, Stefano. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, vol. 47 (2009): 315–72.
- DeScioli, Peter. "The Side-Taking Hypothesis for Moral Judgment." *Current Opinion in Psychology*, vol. 7 (2016): 23–7.
- DeScioli, Peter, and Robert Kurzban. "Morality Is for Choosing Sides." In the *Atlas of Moral Psychology*, edited by Kurt Gray and Jesse Graham. New York: Guilford Press, 2017: 177–85.
- DeScioli, Peter, and Robert Kurzban. "A Solution to the Mysteries of Morality." *Psychological Bulletin*, vol. 139 (2012): 477–96.
- Dewey, John. *Freedom and Culture*. New York: Capricorn Books, 1963.
- Dewey, John. *Liberalism and Social Action*. New York: Perigee Books, 1980.
- Dewey, John. *The Public and Its Problems*. Chicago: Swallow Press, 1927.
- Durkheim, Emile. *The Rules of Sociological Method*, edited by Steven Lukes, translated by W. D. Halls. New York: The Free Press, 1982.
- Durkheim, Emile. *Suicide: A Study in Sociology*, translated by John A. Spaulding and Georges Simpson. New York: Free Press, 1951.
- Durlauf, Steven N. "Complexity, Economics, and Public Policy." *Politics, Philosophy and Economics*, vol. 11 (2012): 41–75.
- The Economist*. "Reforming Cuba." May 16, 2015. <https://www.economist.com/leaders/2015/05/16/be-more-libre>.
- Einav, Liran, and Jonathan Levin. "Economics in the Age of Big Data." *Science*, vol. 346: 6210 (November 2014). doi: 10.1126/science.1243089/.
- Ellickson, Robert C. *Order without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press, 1994.
- Elster, Jon. *Securities against Misrule*. Cambridge: Cambridge University Press, 2013.
- Enquist, M., S. Ghirlanda, A. Jarrick, C.-A. Wachtmeister. "Why Does Human Culture Increase Exponentially?" *Theoretical Population Biology*, vol. 74 (2008) 46–55.

- Ensminger, Jean, and Joseph Henrich, eds. *Experimenting with Social Norms: Fairness and Punishment in a Cross-Cultural Perspective*. New York: Russell Sage Foundation, 2014.
- Estlund, David. "Human Nature and the Limits (if Any) of Political Philosophy." *Philosophy & Public Affairs*, vol. 39 (2011): 207–35.
- Estlund, David. "The Truth in Political Liberalism." In *Truth and Democratic Politics*, edited by Andrew Norris and Jeremy Elkins. Philadelphia: University of Pennsylvania Press, 2012: 251–71.
- Farmer, Charles M., and Adrian K. Lund. "The Effects of Vehicle Redesign on the Risk of Driver Death." *Traffic Injury Prevention*, vol. 16 (2015): 684–90.
- Fehr, Ernst, and Klaus M. Schmidt. "On Inequity Aversion: A Reply to Binmore and Shaked." *Journal of Economic Behavior and Organization*, vol. 73 (2010): 101–8.
- Fehr, Ernst, and Klaus M. Schmidt. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, vol. 114 (1999): 817–68.
- Ferguson, Adam. *An Essay on the History of Civil Society*, edited by Fania Oz-Salzberger. Cambridge: Cambridge University Press, 1995.
- Fischbacher, Urs, Simon Gächter, Ernst Fehr. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, vol. 71 (2001): 397–404.
- Flannery, Kent, and Joyce Marcus. *The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire*. Cambridge, MA: Harvard University Press, 2012.
- Forst, Ranier. "Political Liberty." In *Autonomy and the Challenges to Liberalism*, edited by John Christman and Joel Anderson. Cambridge: Cambridge University Press, 2005: 226–42.
- Freeden, Michael. *Ideologies and Political Theory: A Conceptual Approach*. Oxford: Oxford University Press, 1996.
- Friedman, Daniel. *Morals and Markets: An Evolutionary Account of the Modern World*. New York: Routledge, 2008.
- Garbacz, Christopher. "Do Front-Seat Belt Laws Put Rear-Seat Passengers at Risk?" *Population Research and Policy Review*, vol. 11 (1992): 157–68.
- Garbacz, Christopher. "Estimating Seat Belt Effectiveness with Seat Belt Usage Data from the Centers for Disease Control." *Economics Letters*, vol. 34 (1990): 83–8.
- Gärdenfors, Peter. *Conceptual Spaces*. Cambridge, MA: MIT Press, 2004.
- Gardner, Peter M. "Foragers' Pursuit of Individual Autonomy." *Current Anthropology*, vol. 32 (December 1991): 543–72.
- Gaus, Gerald. "Bernard Bosanquet's Communitarian Defense of Economic Individualism." In *The New Liberalism: Reconciling Liberty and Community*, edited by David Weinstein and Avital Simhony. Cambridge: Cambridge University Press, 2001: 136–58.
- Gaus, Gerald. "A Branch on the Mainline: Hayek's Analysis of Complex Adaptive Systems." *Comos + Taxis*, vol. 7 (2020): 32–40.
- Gaus, Gerald. "The Commonwealth of Bees: On the Impossibility of Justice-through-Ethos." *Social Philosophy & Policy* 33 (2016): 96–121.
- Gaus, Gerald. *Contemporary Theories of Liberalism*. London: Sage, 2003.
- Gaus, Gerald. "Diversity in the Moral Sciences." In *Assessing Austrian Economics*, edited by Daniel J. D'Amico and Adam G. Martin. Bingley, UK: Emerald, 2019: 47–56.
- Gaus, Gerald. "Does Democracy Reveal the Will of the People? Four Takes on Rousseau." *Australasian Journal of Philosophy*, vol. 75 (June 1997): 141–62.

- Gaus, Gerald. "The Evolution, Evaluation and Reform of Social Morality: A Hayekian Analysis." In *Hayek and the Modern Economy*, edited by David Levy and Sandra Peart. New York: Palgrave Macmillan, 2013: 59–88.
- Gaus, Gerald. "The Evolution of Society and Mind: Hayek's System of Ideas." In *The Cambridge Companion to Hayek*, edited by Ed Feser. Cambridge: Cambridge University Press, 2006: 232–58.
- Gaus, Gerald. "Hayekian 'Classical' Liberalism." In the *Routledge Handbook of Libertarianism*, edited by Jason Brennan, Bas van der Vossen, and David Schmidtz. New York: Routledge, 2018: 34–52.
- Gaus, Gerald. "The Idea and Ideal of Capitalism." In *The Oxford Handbook of Business Ethics*, edited by George G. Brenkert and Tom L. Beauchamp. Oxford: Oxford University Press, 2009: 73–99.
- Gaus, Gerald. "In Defense of (Some) Vainglory." In *James M. Buchanan: A Theorist of Political Economy and Social Philosophy*, edited by Richard E. Wagner. New York: Palgrave, 2019: 741–68.
- Gaus, Gerald. "It Can't Be Rational Choice All the Way Down: Comprehensive Hobbesianism and the Origins of the Moral Order." In *Buchanan's Tensions: Reexamining the Political Economy and Philosophy of James M. Buchanan*, edited by Peter J. Boettke, Virgil Henry Storr, and Solomon Stein. Arlington, VA: Mercatus Center, 2018: 117–46.
- Gaus, Gerald. "Liberalism at the End of the Century." *Journal of Political Ideologies*, vol. 5 (2000): 179–99.
- Gaus, Gerald. "A Libertarian Alternative to Liberal Justice." *Criminal Justice Ethics*, vol. 19 (Summer/Fall 2000): 32–43.
- Gaus, Gerald. "Mill's Normative Economics." In *The Blackwell Companion to Mill*, edited by Christopher Macleod and Dale Miller. New York: Wiley-Blackwell, 2017: 488–503.
- Gaus, Gerald. *The Modern Liberal Theory of Man*. New York: Routledge, 2019.
- Gaus, Gerald. "Moral Constitutions." *The Harvard Review of Philosophy*, vol. 19 (2013): 4–22.
- Gaus, Gerald. "The Open Society as a Rule-based Order." *Erasmus Journal for Philosophy and Economics*, vol. 9 (Summer 2016): 1–13.
- Gaus, Gerald. *The Order of Public Reason*. Cambridge: Cambridge University Press, 2011.
- Gaus, Gerald. "Political Philosophy as the Study of Complex Normative Systems." *Cosmos + Taxis*, vol. 5 (January 2018): 62–76.
- Gaus, Gerald. "The Priority of Social Morality." In *Morality, Governance, and Social Institutions: Reflections on Russell Hardin*, edited by Thomas Christiano, Ingrid Creppell and Jack Knight. New York: Palgrave Macmillan, 2018: 23–57.
- Gaus, Gerald. "Private and Public Conscience (Or, Is the Sanctity of Conscience a Liberal Commitment or an Anarchical Fallacy?)" In *Reason, Value, and Respect: Kantian Themes from the Philosophy of Thomas E. Hill, Jr.*, edited by Mark Timmons and Robert Johnson. Oxford: Oxford University Press, 2015: 135–56.
- Gaus, Gerald. "Property." In the *Oxford Handbook of Political Philosophy*, edited by David Estlund. Oxford: Oxford University Press, 2012: 93–112.
- Gaus, Gerald. "Public and Private Interests in Liberal Political Economy, Old and New." In *Public and Private in Social Life*, edited by S. I. Benn and G. F. Gaus. New York: St. Martin's Press, 1983: 183–222.

- Gaus, Gerald. "Retributive Justice and Social Cooperation." In *Retributivism: Essays on Theory and Practice*, edited by Mark D. White. Oxford: Oxford University Press, 2011: 73–90.
- Gaus, Gerald. "The Role of Conservatism in Securing and Maintaining Just Moral Constitutions: Toward a Theory of Complex Normative Systems." In *NOMOS LVI: American Conservatism*, edited by Sanford V. Levinson, Joel Parker, and Melissa S. Williams. New York: New York University Press, 2016: 256–91.
- Gaus, Gerald. "Sectarianism without Perfection? Quong's Political Liberalism." *Philosophy and Public Issues*, vol. 2, no. 2 (Fall 2012): 7–15.
- Gaus, Gerald. "Self-Organizing Moral Systems: Beyond Social Contract Theory," *Politics, Philosophy and Economics*, vol. 17 (2018): 119–47. doi: 10.1177/1470594X17719425.
- Gaus, Gerald. "Social Complexity and Evolved Moral Principles." In *Liberalism, Conservatism, and Hayek's Idea of Spontaneous Order*, edited by Peter McNamara. London: Palgrave Macmillan, 2007: 149–76.
- Gaus, Gerald. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton, NJ: Princeton University Press, 2016.
- Gaus, Gerald. *Value and Justification: The Foundations of Liberal Theory*. Cambridge: Cambridge University Press, 1990.
- Gaus, Gerald. "Why the Conventionalist Needs the Social Contract (and Vice Versa)." *RMM (Rationality, Morality, and Markets)*, vol. 4 (2013): 71–87.
- Gaus, Gerald, and Shaun Nichols. "Moral Learning in the Open Society: The Theory and Practice of Natural Liberty." *Social Philosophy & Policy*, vol. 34 (Summer 2017): 79–101.
- Gauthier, David. *Morals by Agreement*. Oxford: Clarendon Press, 1986.
- Gauthier, David. "Twenty-Five On." *Ethics*, vol. 123 (July 2013): 601–24.
- Gauthier, David. "Why Contractarianism?" In *Contractarianism and Rational Choice*, edited by Peter Vallentyne. Cambridge: Cambridge University Press, 1991: 15–30.
- Gavrilets, Sergey. "High-Dimensional Fitness Landscapes and Speciation" in *Evolution—the Extended Synthesis*, edited by Massimo Pigliucci and Gerd B. Müller. Cambridge, MA: MIT Press, 2010: 45–80.
- Geertz, Clifford. *The Interpretation of Cultures*. New York: Basic Books, 1973.
- Geyer, Robert, and Samir Rihani. *Complexity and Public Policy: A New Approach to 21st Century Politics, Policy and Society*. New York: Routledge, 2010.
- Ghiselin, Michael T. *The Economy of Nature and the Evolution of Sex*. Berkeley: University of California Press, 1974.
- Gillian, Trent. "Weird Things Are Happening to the Market." *Financial Times*, April 5, 2019.
- Gilligan, Michael J., Benjamin J. Pasquale, and Cyrus D. Samii. "Civil War and Social Capital: Behavioral-Game Evidence from Nepal." Electronic copy available at: <http://ssrn.com/abstract=1911969>.
- Gilligan, Michael J., Benjamin J. Pasquale, and Cyrus D. Samii. "Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal." *American Journal of Political Science*, vol. 58 (July 2014): 604–19.
- Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr, eds. *Moral Sentiments and Material Interests: The Foundation of Cooperation in Economic Life*. Cambridge, MA: MIT Press, 2005.

- Gintis, Herbert, Carel van Schaik, and Christopher Boehm. "Zoon Politikon: The Evolutionary Origins of Human Political Systems." *Current Anthropology*, vol. 56 (June 2015): 327–53.
- Goldberg, Jonah. *Suicide of the West: How the Rebirth of Tribalism, Populism, Nationalism and Identity Politics Is Destroying American Democracy*. New York: Crown, 2018.
- Goldfarb, Robert S. "Now You See it, Now You Don't: Emerging Contrary Results in Economics." *Journal of Economic Methodology*, vol. 4 (1997): 221–44.
- Gowdy, John, Mariana Mazzucato, Jeroen C. J. M. van den Bergh, Sander E. van der Leeuw, and David S. Wilson. "Shaping the Evolution of Complex Societies." In *Complexity and Evolution*, edited by David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, 2016: 327–50.
- Graham, Ottis L., Jr. *Toward a Planned Society: From Roosevelt to Nixon*. Oxford: Oxford University Press, 1976.
- Gray, John. *Hayek on Liberty*. Oxford: Blackwell, 1984.
- Green, Sara. "Can Biological Complexity Be Reverse Engineered?" *Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 53 (2015): 73–83.
- Greene, Joshua. *Moral Tribes: Emotion, Reason and the Gap between Us and Them*. New York: Penguin, 2013.
- Gressman, Philip T., and Jennifer R. Peck. "Simulating COVID-19 in a University Environment." Accessed June 28, 2020, at <https://arxiv.org/pdf/2006.03175.pdf>.
- Gurven, Michael. "To Give and to Give Not: The Behavioral Ecology of Human Food Transfers." *Behavioral and Brain Sciences*, vol. 27 (2004): 543–83.
- Gurven, Michael, Wesley Allen-Arave, Kim Hill, and A. Magdalena Hurtado. "Reservation Food Sharing among the Ache of Paraguay." *Human Nature*, vol. 12 (2001): 273–97.
- Gurven, Michael, Wesley Allen-Arave, Kim Hill, Magdalena Hurtado. "It's a Wonderful Life': Signaling Generosity among the Ache of Paraguay." *Evolution and Human Behavior*, vol. 21 (2000): 263–82.
- Haidt, Jonathan. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage, 2013.
- Hamilton, W. D. "The Genetical Evolution of Social Behaviour: I." *Journal of Theoretical Biology*, vol. 7 (1964): 1–16.
- Hamlin, John. "Feasibility Four Ways." *Social Philosophy & Policy*, vol. 34 (2017) 209–31.
- Hardin, Russell. *David Hume: Moral and Political Theorist*. Oxford: Oxford University Press, 2007.
- Hardin, Russell. "From Bodo Ethics to Distributive Justice." *Ethical Theory and Moral Practice* vol. 2 (1999): 399–413.
- Hardin, Russell. "The Priority of Social Order." *Rationality and Society*, vol. 25 (2013): 407–21.
- Harvey, A. C., and J. Durbin. "The Effects of Seat Belt Legislation on British Road Deaths." *Journal of the Royal Statistical Society, Series A*, vol. 149 (1986): 187–210.
- Hausman, Daniel M., and Michael S. Macpherson. *Economic Analysis, Moral Philosophy and Public Policy*, 2nd ed. Cambridge: Cambridge University Press, 2006.
- Hayek, F. A. *The Constitution of Liberty*. Chicago: University of Chicago Press, 1960.
- Hayek, F. A. *The Counter-Revolution of Science*. New York: Free Press, 1955.
- Hayek, F. A. "Degrees of Explanation." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 195–212.

- Hayek, F. A. "Dr. Bernard Mandeville." In his *New Studies in Philosophy, Politics Economics, and the History of Ideas*. New York: Routledge, 1978: 249–69.
- Hayek, F. A. "Economics and Knowledge." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 57–77.
- Hayek, F. A. "Economists and Philosophers." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press 2014: 427–43.
- Hayek, F. A. "Engineers and Planners." In *Studies on the Abuse and Decline of Reason*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2010: 56–66.
- Hayek, F. A. *The Fatal Conceit: The Errors of Socialism*, edited by W. W. Bartley III. Chicago: University of Chicago Press, 1988.
- Hayek, F.A. "Individualism True and False." In *Studies on the Abuse and Decline of Reason*, edited by Bruce Caldwell. Chicago: University of Chicago Press 2010: 46–74.
- Hayek, F. A. "Kinds of Rationalism." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 39–54.
- Hayek, F. A. *The Mirage of Social Justice*. Chicago: University of Chicago Press, 1976.
- Hayek, F. A. "Notes on the Evolution of Systems of Rules of Conduct." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 278–92.
- Hayek, F. A. "The Political Ideal of the Rule of Law." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 119–94.
- Hayek, F. A. *The Political Order of a Free People*. Chicago: University of Chicago Press, 1979.
- Hayek, F. A. "The Pretence of Knowledge." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 362–72.
- Hayek, F. A. "The Results of Human Action but Not of Design." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 293–303.
- Hayek, F. A. *The Road to Serfdom*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2007.
- Hayek, F.A. "Rules, Perception and Intelligibility." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, 2014: 232–54.
- Hayek, F. A. *Rules and Order*, vol. 1 of *Law, Legislation and Liberty*. Chicago: University of Chicago Press, 1973.
- Hayek, F. A. *The Sensory Order*. Chicago: University of Chicago Press, 1952.
- Hayek, F. A. "The Theory of Complex Phenomena." In *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press 2014: 257–77.
- Hayek, F. A. "The Use of Knowledge in Society." *American Economic Review*, vol. 35 (September 1945): 519–30.
- Head, John. *Public Goods and Public Welfare*. Durham, NC: Duke University Press, 1974.
- Henrich, Joseph. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press, 2016.
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. "The Weirdest People in the World?" *Behavioral and Brain Sciences*, vol. 33 (2010): 61–135.
- Henrich, Joseph, and Natalie Henrich. "The Evolution of Cultural Adaptations: Fijian Food Taboos Protect against Dangerous Marine Toxins." *Proceedings of the Royal Society B*, vol. 277 (2010): 3715–3724.
- Henrich, Joseph, and Natalie Smith. "Comparative Experimental Evidence from Machiguenga, Mapuche, Huinca, and American Populations." In *Foundations of Human*

- Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, edited by Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. Oxford: Oxford University Press, 2004: 125–67. doi: 10.1093/0199262055.003.0005.
- Henrich, Natalie, and Joseph Henrich. *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. New York: Oxford University Press, 2007.
- Herrmann, Esther, Josep Call, Maria Victoria Hernández-Lloreda, Brian Hare, and Michael Tomasello. “Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis.” *Science*, vol. 317, no. 5843 (2007): 1360–6.
- Herrmann, Esther, Josep Call, Maria Victoria Hernández-Lloreda, Brian Hare, and Michael Tomasello. “The Structure of Individual Differences in the Cognitive Abilities of Children and Chimpanzees.” *Psychological Science*, vol. 21 (January 2010): 102–10.
- Hesse, Mary B. *Models and Analogies in Science*. Notre Dame, IN: Notre Dame University Press, 1960.
- Hehir, Brid. *The Emperor’s New Clothes? The Dangers of the Anti-FGM Campaign*. 2014. http://www.battleofideas.org.uk/BiP_2014/BiP_FGM_Brid_Hehir.pdf.
- Hill, Thomas, Jr. “Four Conceptions of Conscience.” in *NOMOS XL: Integrity and Conscience*, edited by Ian Shapiro and Robert Adams. New York: New York University Press, 1998: 13–52.
- Hodgson, Geoffrey M. 1993. *Economics and Evolution: Bringing Life Back into Economics*. Ann Arbor: University of Michigan Press, 1993.
- Hoitink, Yvette. “Why the Little Dutch Boy Never Put His Finger in the Dike.” *Dutch Genealogy*, May 3, 2020. <https://www.dutchgenealogy.nl/why-the-little-dutch-boy-never-put-his-finger-in-the-dike/>.
- Holland, John H. *Complexity*. Oxford: Oxford University Press, 2014.
- Holland, John H. *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. Cambridge, MA: MIT Press, 2012.
- Hong, Lu, and Scott E. Page. “Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 46 (November 16, 2004): 16385–9.
- Hong, Lu, and Scott E. Page. “Problem Solving by Heterogeneous Agents.” *Journal of Economic Theory*, vol. 97 (2001): 123–63.
- Hopfensitz, Astrid, and Ernesto Reuben. “The Importance of Emotions for the Effectiveness of Social Punishment.” *The Economic Journal*, vol. 119 (October 2009): 1534–59.
- Horner, Victoria, and Andrew Whiten. “Causal Knowledge and Imitation/Emulation in Chimpanzees (*Pan Troglodytes*) and Children (*Homo Sapiens*).” *Animal Cognition*, vol. 8 (2005): 164–81.
- Houston, David J., and Lilliard E. Richardson, Jr. “Reducing Traffic Fatalities in the American States by Upgrading Seat Belt Use Laws to Primary Enforcement.” *Journal of Policy Analysis and Management*, vol. 25 (2006): 645–59.
- Houston, David J., and Lilliard E. Richardson, Jr. “Traffic Safety and the Switch to a Primary Seat Belt Law: The California Experience.” *Accident Analysis and Prevention* 34 (2002): 743–51.
- Houston, David J., Lilliard E. Richardson, Jr., and Grant W. Neely, “Legislating Traffic Safety: A Pooled Time Series Analysis.” *Social Science Quarterly*, vol. 76 (1995): 330–1.

- Hume, David. *A Treatise of Human Nature*, 2nd ed., edited by P. H. Nidditch. Oxford: Clarendon Press, 1978.
- Irwin, Alan. "Technical Expertise and Risk Conflict: An Institutional Study of the British Compulsory Seat Belt Debate." *Policy Sciences*, vol. 20 (1987): 339–64.
- Ismael, J. T. *How Physics Makes Us Free*. Oxford: Oxford University Press, 2016.
- Ismael, J. T. "Self-Organization and Self-Governance." *Philosophy of the Social Sciences*, vol. 41 (2011): 327–51.
- Jacobs, Jane. *The Death and Life of Great American Cities*. New York: Vintage, 1992.
- Jenness, Diamond. *The Life of the Copper Eskimos*, Volume 12 of *Report of the Canadian Arctic Expedition, 1913–18*. Ottawa: F. A. Ackland, 1922.
- Jervis, Robert. "Complexity and the Analysis of Political and Social Life." *Political Science Quarterly*, vol. 112 (1997–98): 569–93.
- Johnson, James. "Review Essay—Formal Models in Political Science: Conceptual, Not Empirical." *Journal of Politics*, vol. 81 (2018): 6–10. <http://dx.doi.org/10.1086/700590>.
- Johnson, Neil. *Simply Complexity: A Clear Guide to Complexity Theory*. London: Oneworld, 2007.
- Joyce, Richard. *The Evolution of Morality*. Cambridge, MA: MIT Press, 2007.
- Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.
- Kahneman, Daniel, and Amos Tversky. "On the Psychology of Prediction." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky. Cambridge: Cambridge University Press, 1982: 48–68.
- Kant, Immanuel. *Foundations of the Metaphysics of Morals*, translated by Lewis White Beck. Indianapolis: Bobbs-Merrill, 1959.
- Kant, Immanuel. *The Metaphysical Elements of Justice*, 2nd ed., edited and translated by John Ladd. Indianapolis: Hackett, 1999.
- Kaplan, Benjamin. *Divided by Faith: Religious Conflict and the Practice of Toleration in Early Modern Europe*. Cambridge, MA: Belknap Press of Harvard University Press, 2007.
- Kaplan, Hillard, and Michael Gurvan. "The Natural History of Food Sharing: A Review and a New Multi-Individual Approach to the Negotiation of Norms." In *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, edited by Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr. Cambridge, MA: MIT Press, 2005: 75–113.
- Kauffman, Stuart A. *The Origins of Order*. New York: Oxford University Press, 1993.
- Kauffman, Stuart A. *A World beyond Physics: The Emergence of Evolution and Life*. Oxford: Oxford University Press, 2017.
- Kavka, Gregory. "The Reconciliation Project." In *Morality, Reason and Truth: New Essays on the Foundations of Ethics*, edited by David Copp and David Zimmerman. Totowa, NJ: Rowman and Allenheld, 1985: 297–319.
- Kennedy, Robert. *Thirteen Days: A Memoir of the Cuban Missile Crisis*. New York: W. W. Norton, 1969.
- Keynes, John Maynard. "Am I a Liberal?" In his *Essays in Persuasion*. London: Macmillan, 1972: 272–94.
- Keynes, John Maynard. "The End of Laissez-Faire." In his *Essays in Persuasion*. London: Macmillan, 1972: 312–22.
- Keynes, John Maynard. *The General Theory of Employment, Interest and Money*. London: Macmillan, 1974.

- Khalil, Elias L. "Friedrich Hayek's Darwinian Theory of Evolution of Institutions: Two Problems." *Australian Economic Papers*, vol. 35 (June 1996): 183–201.
- Kirman, Alan, and Rajiv Sethi. "Disequilibrium Adjustment and Economic Outcomes." In *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, 2016: 15–30.
- Kitcher, Philip. "The Division of Cognitive Labor." *The Journal of Philosophy*, vol. 87 (1990): 5–22.
- Kitcher, Philip. *The Ethical Project*. Cambridge, MA: Harvard University Press, 2011.
- Knight, Frank. *Risk, Uncertainty, and Profit*. New York: Augustus M. Kelly, 1964.
- Knight, Jack, and James Johnson. *The Priority of Democracy: Political Consequences of Pragmatism*. Princeton, NJ: Princeton University Press, 2011.
- Kuhn, Thomas S. "Second Thoughts on Paradigms." In his *The Essential Tension: Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press, 1977: 293–319.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: University of Chicago Press, 1970.
- Kukathas, Chandran. *Hayek and Modern Liberalism*. Oxford: Clarendon Press, 1990.
- Kukathas, Chandran. *The Liberal Archipelago: A Theory of Diversity and Freedom*. Oxford: Oxford University Press 2003.
- Laland, Kevin N. *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press, 2017.
- Laland, Kevin N., and Luke Rendell. "Cultural Memory." *Current Biology*, vol. 23 (2013): R736–40.
- Landemore, Hélène. *Democratic Reason*. Princeton, NJ: Princeton University Press, 2013.
- Landemore, Hélène, and Scott E. Page, "Deliberation and Disagreement: Problem Solving, Prediction, and Positive Dissensus," *Politics, Philosophy & Economics*, vol. 14 (2014): 1–26.
- Lane, Ruth. *The Complexity of Self-Government: Politics from the Bottom-Up*. Cambridge: Cambridge University Press, 2017.
- Ledyard, John O. "Public Goods: A Survey of Experimental Research." In *Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press, 1995: 111–93.
- Lee, Richard Borshay. *The !Kung San: Men, Women and Work in a Foraging Society*. Cambridge: Cambridge University Press, 1979.
- Levinthal, Daniel A., and Massimo Warglien. "Landscape Design: Designing for Local Action in Complex Worlds." *Organization Science*, vol. 10 (special issue: Application of Complexity Theory to Organization Science) (May–June 1999): 342–357.
- Levy, David M., and Sandra J. Peart. *Escape from Democracy: The Role of Experts and the Public in Economic Policy*. Cambridge: Cambridge University Press, 2017.
- Levy, Jacob T. *Rationalism, Pluralism and Freedom*. Oxford: Oxford University Press, 2015.
- Lewis, Paul. "An Analytical Core for Sociology: A Complex, Hayekian Analysis." *Review of Behavioral Economics*, vol. 2 (2015): 123–36.
- Lewis, Paul. "The Emergence of 'Emergence' in the Work of F. A. Hayek." *History of Political Economy*, vol. 48 (2016): 111–50.
- Lewis, Paul. "Emergent Properties in the Work of Friedrich Hayek." *Journal of Economic Behavior and Organization*, vol. 82 (2012): 368–78.

- Lewis, Paul. "Notions of Order and Process in Hayek: The Significance of Emergence." *Cambridge Journal of Economics*, vol. 39 (2015): 1167–90.
- Lewis, Paul. "Ontology and the History of Economic Thought: The Case of Reductionism in the Work of F. A. Hayek." *Cambridge Journal of Economics*, vol. 41 (2017): 1343–65.
- Lewis, Paul. "The Ostroms and Hayek as Theorists of Complex Adaptive Systems: Commonality and Complementarity." In *The Austrian and Bloomington Schools of Political Economy*, edited by Paul Dragos Aligica, Paul Lewis, and Virgil Henry Storr. Bingley, UK: Emerald, 2017: 35–55.
- Lewis, Paul. "Purposeful Behaviour, Expectations, and the Mirage of Social Justice: The Influence of Cybernetics on the Thought of F. A. Hayek (2016)." Electronic copy available at: <http://ssrn.com/abstract=2790169>.
- Lewis, Paul. "Systems, Levels of Organisation and Structural Properties: The Influence of Ludwig von Bertalanffy on the Development of F. A. Hayek's Work on Complexity." In *Research in the History of Economic Thought and Methodology*, vol. 34A, edited by Luca Fiorito, S. Scheall and C. Suprinyak. Bingley: Emerald, 2016: 125–59.
- Libecap, Gary D. *Contracting for Property Rights*. Cambridge: Cambridge University Press, 1989.
- Lijphart, Arend. *Democracy in Plural Societies: A Comparative Analysis*. New Haven, CT: Yale University Press, 1977.
- Lindblom, Charles E. "The Science of 'Muddling' Through." *Public Administration Review*, vol. 19, no. 2: 79–88.
- Locke, John. *Some Considerations of the Consequences of the Lowering of Interest, and Raising the Value of Money*. In *The History of Economic Thought*, edited by Steven G. Medema and Warren J. Samuels. New York: Routledge, 2013: 63–77.
- López-Pérez, Raúl. "Aversion to Norm-Breaking: A Model." *Games and Economic Behavior*, vol. 64 (2008): 237–67.
- Lorenz, Edward. *The Essence of Chaos*. Seattle: University of Washington Press, 1993.
- Lorenz, Konrad. *On Aggression*. New York: Routledge, 2002.
- Lycett, Stephen J., Mark Collard, and William C. McGrew. "Phylogenetic Analyses of Behavior Support Existence of Culture Among Wild Chimpanzees." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104 (November 6, 2007): 17588–92.
- Macedo, Stephen. "Why Public Reason? Citizens' Reasons and the Constitution of the Public Sphere." Available at: <http://ssrn.com/abstract=1664085>.
- Mack, Eric. "Hayek on Justice and the Order of Actions." In *The Cambridge Companion to Hayek*, edited by Ed Feser. Cambridge: Cambridge University Press, 2006: 259–86.
- Mack, Eric. "What Is Seen and What Is Not Seen: Hayekian Extensions of Bastiat's Insight." *Georgetown Journal of Law and Public Policy*, vol. 15, special issue (2017): 811–34.
- Mackie, Gerry. "Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience." In *Cultural Agents Reloaded: The Legacy of Antanas Mockus*, edited by Carlo Tognato. Cambridge, MA: The Cultural Agents Initiative at Harvard University Press, 2017: 313–35.
- Mackie, J. L. *Ethics: Inventing Right and Wrong*. Harmondsworth, UK: Penguin, 1990.
- Malthus, Thomas. *An Essay on the Principle of Population*. London: Printed for J. Johnson, in St. Paul's Church-Yard, 1798.

- Mark, Noah P. "Status Organizes Cooperation: An Evolutionary Theory of Status and Social Order." *American Journal of Sociology*, vol. 23 (2018): 1601–34.
- Martin, Christopher Flynn. Rahul Bhui, Peter Bossaerts, Tetsuro Matsuzawa, and Colin Camerer. "Chimpanzee Choice Rates in Competitive Games Match Equilibrium Game Theory Predictions." *Scientific Reports*, vol. 4 (2014): 5182. doi: 10.1038/srep05182.
- Maynard Smith, John. "Group Selection and Kin Selection." *Nature*, vol. 201 (1964): 1145–47.
- McCloskey, Deirdre Hansen. *Bourgeois Equality*. Chicago: University of Chicago Press, 2016.
- McGrew, W. C., and Anna T. Feistner. "Two Nonhuman Primate Models for the Evolution of Human Food Sharing: Chimpanzees and Callitrichids." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by Jerome H. Barkow, Leda Cosmides, and John Tooby. New York: Oxford University Press, 1992: 229–43.
- McKelvey, Bill. "Avoiding Complexity Catastrophe in Coevolutionary Pockets: Strategies for Rugged Landscapes." *Organization Science*, vol. 10 (special issue: Application of Complexity Theory to Organization Science) (May–June 1999): 294–321.
- Melis, Alicia P., Brian Hare, and Michael Tomasello. "Do Chimpanzees Reciprocate Received Favours?" *Animal Behaviour*, vol. 76 (2008): 951–62.
- Mercier, Hugo, and Dan Sperber. *The Enigma of Reason*. Cambridge, MA: Harvard University Press, 2017.
- Mesoudi, Alex. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press, 2011.
- Mikhail, John. *The Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press, 2010.
- Miler, John H. *A Crude Look at the Whole: The Science of Complex Systems in Business, Life, and Society*. New York: Basic Books, 2015.
- Mill, John Stuart. "Civilisation." In *The Collected Works of John Stuart Mill*, edited by J. M. Robson, vol. 18. Toronto: University of Toronto Press, 1977: 117–48.
- Mill, John Stuart. *On Liberty*. In *The Collected Works of John Stuart Mill*, edited by J. M. Robson, vol. 18. Toronto: University of Toronto Press, 1977: 213–310.
- Mill, John Stuart. *The Principles of Political Economy with Some of Their Applications to Social Philosophy*. In *The Collected Works of John Stuart Mill*, edited by J. M. Robson, vols. II and III. Toronto: University of Toronto Press, 1965.
- Mill, John Stuart. *A System of Logic*. In *The Collected Works of John Stuart Mill*, edited by J. M. Robson, vol. 7. Indianapolis: Liberty Fund, 2006.
- Mill, John Stuart. "The Subjection of Women." In *The Collected Works of John Stuart Mill*, edited by J. M. Robson, vol. 21. Toronto: University of Toronto Press, 1984: 259–341.
- Miller, David. "The Fatalistic Conceit." *Critical Review*, vol. 3 (Spring 1989): 310–23.
- Miller, Geoffrey. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. New York: Anchor Books, 2001.
- Miller, John H. *A Crude Look at the Whole: The Science of Complex Systems in Business, Life and Society*. New York: Basic Books, 2015.
- Miller, John H., and Scott E. Page. 2007. *Complex Adaptive Systems*. Princeton, NJ: Princeton University Press.

- Mitchell, Melanie. *Complexity*. Oxford: Oxford University Press, 2009.
- Mockus, Antanas. "Building 'Citizenship Culture' in Bogotá." *Journal of International Affairs*, vol. 65 (Spring–Summer 2012): 143–6.
- Moehler, Michael. *Minimal Morality: A Multilevel Social Contract Theory*. New York: Oxford University Press, 2018.
- Moore, Jim. "Female Transfer in Primates." *International Journal of Primatology*, vol. 5 (1984): 537–89.
- Mueller, Dennis C. *Public Choice III*. Cambridge: Cambridge University Press, 2003.
- Muldoon, Ryan. "Diversity and the Division of Cognitive Labor." *Philosophy Compass*, vol. 8 (2013): 117–25.
- Muldoon, Ryan. "Perspectives, Norms and Agency," *Social Philosophy & Policy*, vol. 34 (2017): 260–76.
- Muldoon, Ryan. *Social Contract Theory for a Diverse World*. New York: Routledge, 2016.
- Muldoon, Ryan. "Understanding Norms and Changing Them." *Social Philosophy and Policy*, vol. 35 (2018): 128–48.
- Muldoon, Ryan, Michael Borgida, and Michael Cuffaro. "The Conditions of Tolerance." *Politics, Philosophy & Economics*, vol. 11 (2011): 322–44.
- Muldoon, Ryan, Chiara Liscandra, Mark Colyvan, Carlo Martini, Giacomo Sillari, and Jan Sprenger. "Disagreement behind the Veil of Ignorance." *Philosophical Studies*, vol. 170 (2014): 377–94.
- Müller, Julian. *Polycentric Democracy: Making Use of Diversity*. New York: Routledge, 2019.
- National Academies of Sciences, Engineering, and Medicine. *The Economic and Fiscal Consequences of Immigration*. Washington, DC: The National Academies Press, 2017. <https://doi.org/10.17226/23550>.
- Nichols, Shaun, and Gerald Gaus. "Unspoken Rules: Resolving Underdetermination with Closure Principles." *Cognitive Science*, vol. 42 (2018): 2735–56.
- Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening. "Normative Conflict and Feuds: The Limits of Self-Enforcement." *Journal of Public Economics*, vol. 96 (2012): 797–807.
- North, Douglass C. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press, 1990.
- Nowak, Martin A., and Roger Highfield. *Super Cooperators: Altruism, Evolution, and Why We Need Each Other*. New York: Free Press, 2011.
- Nowak, Martin A., and Karl Sigmund. "Evolution of Indirect Reciprocity." *Nature*, vol. 437 (October 27, 2005): 1291–98.
- Nowak, Martin A., C. E. Tarnita, and E. O. Wilson. "The Evolution of Eusociality." *Nature*, vol. 466 (2010): 1057–62.
- Nozick, Robert. *Anarchy, State and Utopia*. New York: Basic Books, 1974.
- Nunnar-Winkler, Gertrude, and Beate Sodian. "Children's Understanding of Moral Emotions." *Child Development*, vol. 59 (October 1988): 1323–38.
- O'Brien, D. P. *The Classical Economists Revisited*. Princeton, NJ: Princeton University Press, 2004.
- O'Connor, Cailin. *The Origins of Unfairness*. New York: Oxford University Press, 2019.
- O'Hear, Anthony. *Beyond Evolution: Human Nature and the Limits of Evolutionary Explanation*. Oxford: Oxford University Press, 1997.

- Oakeshott, Michael. *On Human Conduct*. Oxford: Oxford University Press, 1975.
- Oakeshott, Michael. "Rationalism in Politics." In his *Rationalism in Politics and Other Essays*, expanded ed. Indianapolis: Liberty Press, 1991: 5–42.
- Okasha, Samir. *Evolution and Levels of Selection*. Oxford: Oxford University Press, 2006.
- Ostrom, Elinor. "An Agenda for the Study of Institutions." In *Choice, Rules and Collective Action: The Ostroms on the Study of Institutions and Governance*, edited by Fillippo Sabetti and Paul Dragos Aligica. Essex, UK: ECPR Press, 2014: 97–119.
- Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press, 1990.
- Ostrom, Elinor, and Vincent Ostrom. "The Quest for Meaning in Public Choice." In *Choice, Rules and Collective Action: The Ostroms on the Study of Institutions and Governance*, edited by Fillippo Sabetti and Paul Dragos Aligica. Essex: ECPR Press, 2014: 61–86.
- Ostrom, Vincent. *The Intellectual Crisis in American Public Administration*, 3rd ed. Tuscaloosa: University of Alabama Press, 2008.
- Ostrom, Vincent. "Polycentricity: The Structural Basis of Self-Governing Systems." In *Choice, Rules and Collective Action: The Ostroms on the Study of Institutions and Governance*, edited by Fillippo Sabetti and Paul Dragos Aligica. Essex: ECPR Press, 2014: 45–60.
- Ostrom, Vincent, and Elinor Ostrom. "Public Goods and Public Choice." In *Polycentricity and Local Public Economics*, edited by Michael D. McGinnis. Ann Arbor: University of Michigan Press, 2002: 75–103.
- Ostrom, Vincent, Charles M. Tiebout, and Robert Warren. "The Organization of Government in Metropolitan Areas: A Theoretical Enquiry." *American Political Science Review*, vol. 55 (December 1961): 831–42.
- Over, Harriet, and Malinda Carpenter. "The Social Side of Imitation." *Child Development Perspectives*, vol. 7 (2013): 6–11.
- Page, Scott E. *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy*. Princeton, NJ: Princeton University Press, 2017.
- Page, Scott E. *Diversity and Complexity*. Princeton, NJ: Princeton University Press, 2011.
- Page, Scott E. "The Imperative of Complexity." *Cosmos + Taxis*, vol. 5 (January 2018): 4–12.
- Page, Scott E. *The Model Thinker*. New York: Basic Books, 2018.
- Page, Scott E. "Not Half Bad: A Modest Criterion for Inclusion." In *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, 2016: 319–26.
- Page, Scott E., and David W. Richardson. "Walsh Functions, Schema Variance, and Deception." *Complex Systems*, vol. 6 (1992): 125–35.
- Paluck, Elizabeth Levy, and Hana Shephard. "The Salience of Social Referents: A Field Experiment on Collective Norms and Harassment Behavior in a School Social Network." *Journal of Personality and Social Psychology*, vol. 103 (2012): 899.
- Parsons, Talcott. *Social Action: A Study in Social Theory with Special Reference to a Group of Recent European Writers*, 2nd ed. Glencoe, IL: Free Press, 1949.
- Paul, Ellen Frankel. "Liberalism, Unintended Orders and Evolutionism." *Political Studies*, vol. 36 (1988): 251–72.
- Peltzman, Sam. "The Effects of Automobile Safety Regulations." *Journal of Political Economy*, vol. 83 (August 1975): 677–726.
- Pert, Sandra J., and David M. Levy. "Discussion, Construction and Evolution: Mill, Buchanan and Hayek on Constitutional Order." *Constitutional Political Economy*, vol. 19 (2008): 3–18.

- Peters, Uwe, Nathan Honeycutt, Andreas De Block, and Lee Jussim. "Ideological Diversity, Hostility, and Discrimination in Philosophy." *Philosophical Psychology*, vol. 33, no. 4 (2020): 511–548. doi: 10.1080/09515089.2020.1743257.
- Pettit, Philip. *On the People's Terms*. Cambridge: Cambridge University Press, 2012.
- Pettit, Philip. *Republicanism*. Oxford: Clarendon Press, 1997.
- Pinker, Steven. *How the Mind Works*. New York: W. W. Norton, 1999.
- Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin, 2002.
- Platteau, Jean-Philippe. *Institutions, Social Norms, and Economic Development*. New York: Routledge, 2000.
- Popper, Karl. *After the Open Society: Selected Social and Political Writings*, edited by Jeremy Shearmur and Piers Norris Turner. New York: Routledge, 2008.
- Popper, Karl. *The Open Society and Its Enemies*, 4th ed., two vols. London: Routledge and Kegan Paul, 1962.
- Popper, Karl. *Unended Quest: An Intellectual Autobiography*. London: Fontana/Collins, 1982.
- Preuschoft, S., and Jan A. R. A. M. van Hooff. "Homologizing Primate Facial Displays: A Critical Review of Methods." *Folia Primatologica*, vol. 65 (1996): 121–37.
- Prinz, Jesse. *The Emotional Construction of Morals*. Oxford: Oxford University Press, 2007.
- Quong, Jonathan. *Liberalism without Perfectionism*. Oxford: Oxford University Press, 2011.
- Rawls, John. "The Independence of Moral Theory," In *John Rawls: Collected Papers*, ed. Samuel Freeman. Cambridge, MA: Harvard University Press, 1999: 286–302.
- Rawls, John. *Justice as Fairness: A Restatement*, edited by Erin Kelly. Cambridge: Belknap Press of Harvard University Press, 2001.
- Rawls, John. "Justice as Fairness." In *John Rawls: Collected Papers*, edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 1999: 347–72.
- Rawls, John. "Kantian Constructivism." In *John Rawls: Collected Papers*, edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 1999: 303–58.
- Rawls, John. "Outline of a Decision Procedure for Ethics." In *John Rawls: Collected Papers*, edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 1999: 1–19.
- Rawls, John. *Political Liberalism*, expanded ed. New York: Columbia University Press, 2005.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
- Rawls, John. *A Theory of Justice*, rev. ed. Cambridge, MA: Belknap Press of Harvard University Press, 1999.
- Rendell, L., R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland. "Why Copy Others? Insights from the Social Learning Strategies Tournament." *Science*, new series, vol. 328 (April 9, 2010): 208–213.
- Rhoads, Steven. *The Economist's View of the World*. Cambridge: Cambridge University Press, 1985.
- Richerson, Peter J., and Robert Boyd. "The Evolution of Free Enterprise Values." In *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak. Princeton, NJ: Princeton University Press, 2008: 107–41.
- Richerson, Peter J., and Robert Boyd. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press, 2005.
- Richerson, Peter J., and Robert Boyd. "Rethinking Paleoanthropology: A World Queerer Than We Supposed." In *Evolution of Mind*, edited by Gary Hatfield and Holly Pittman. Philadelphia: Pennsylvania Museum Conference Series, 2013: 263–302.

- Ridley, Matt. *The Evolution of Everything: How New Ideas Emerge*. New York: Harper, 2015.
- Robbins, Lionel. *A History of Economic Thought*, edited by Steven G. Medema and Warren J. Samuels. Princeton, NJ: Princeton University Press, 1998.
- Robbins, Lionel. *Political Economy Past and Present*. London: Macmillan, 1977.
- Robbins, Lionel. *The Theory of Economic Policy in Classical Political Economy*. London: Macmillan, 1952.
- Robinson, Paul. "Why Does the Criminal Law Care What the Layperson Thinks Is Just?" *Virginia Law Review*, vol. 86 (2000): 1839–69.
- Rogers, Alan R. "Does Biology Constrain Culture?" *American Anthropologist*, vol. 90 (December 1988): 819–31.
- Rose, David C. *The Moral Foundations of Economic Behavior*. New York: Oxford University Press, 2011.
- Rose, David. "On the Evolution of Ethics, Rationality, and Economic Behavior." In *The Oxford Handbook of Ethics and Economics*, edited by Mark D. White. New York: Oxford University Press, 2019: 16–78.
- Rose, David C. *Why Culture Matters Most*. Oxford: Oxford University Press, 2019.
- Rosenberg, Alexander. "On the Very Idea of Ideal Theory in Political Philosophy." *Social Philosophy and Policy*, vol. 33 (2016): 55–75.
- Rosenberg, Shawn, Dana Ward, and Stephen Chilton. *Political Reasoning and Cognition: A Piagetian View*. Durham, NC: Duke University Press, 1988.
- Rosser, J. Barkley, Jr. "On the Complexities of Complex Economics Dynamics." *The Journal of Economic Perspectives* 13 (Autumn 1999): 169–92.
- Rostow, W. W. *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge: Cambridge University Press, 1960.
- Rousseau, Jean-Jacques. *Discourse on the Origin and the Foundation of Inequality among Men*. In *The Discourses and Other Early Political Writings*, edited by Victor Gourevitch. Cambridge: Cambridge University Press, 1997: 111–222.
- Rousseau, Jean-Jacques. *Of the Social Contract, Or Principles of Political Right*. In *The Social Contract and Other Later Political Writings*, edited by Victor Gourevitch. Cambridge: Cambridge University Press, 1997: 39–152.
- Rubinstein, Ariel. *Economic Fables*. Cambridge: Open Book, 2012.
- Ryan, David L., and Guy A. Bridgeman. "Judging the Roles of Legislation, Education, and Offsetting Behaviour in Seat Belt Use: A Survey and New Evidence from Alberta." *Canadian Public Policy*, vol. 18 (1992): 27–46.
- Saari, Donald. "Mathematical Complexity of Simple Economics." *Notices of the AMA*, vol. 42, no. 2 (1995): 222–231.
- Sabel, Charles F., and Jonathan Zeitlin. "Experimentalist Governance." In *The Oxford Handbook of Governance*, edited by David Levi-Faur. Oxford: Oxford University Press, 2012: 169–84.
- Sabel, Charles F., and Jonathan Zeitlin. "Learning from Difference: The New Architecture of Experimentalist Governance in the EU." *European Law Journal*, vol. 14 (May 2008): 271–327.
- Sarkissian, Hagop, John Park, David Tien, Jennifer Cole Wright, and Joshua Knobe. "Folk Moral Relativism." *Mind & Language*, vol. 26 (September 2011): 482–505.
- Scheall, Scott. *F. A. Hayek and the Epistemology of Politics: The Curious Task of Economics*. New York: Routledge, 2020.

- Schelling, Thomas. *Motives and Mircobehavior*. New York: Norton, 1978.
- Schneider, Volker. "Governance and Complexity." In *The Oxford Handbook of Governance*, edited by David Levi-Faur. Oxford: Oxford University Press, 2012: 129–40.
- Schulz, Jonathan F., Duman Barahmi-Rad, Jonathan Beauchamp, and Joseph Henrich. "The Origins of WEIRD Psychology." *PsyArXiv* (June 22, 2018). doi: 10.31234/osf.io/d6qhu.
- Schumpeter, Joseph A. *Capitalism, Socialism and Democracy*, 3rd ed. London: Allen and Unwin, 1950.
- Schwab, David, and Elinor Ostrom. "The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies." In *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak. Princeton, NJ: Princeton University Press, 2008: 204–27.
- Scott, James C. *Against the Grain: A Deep History of the Earliest States*. New Haven, CT: Yale University Press, 2017.
- Sen, Amartya. *Collective Choice and Social Welfare*, expanded ed. Cambridge, MA: Harvard University Press, 2017.
- Sen, Amartya. *The Idea of Justice*. Cambridge, MA: Harvard University Press, 2009.
- Sen, Amartya. "Rational Fools." In his *Choice, Welfare and Measurement*. Cambridge, MA: Harvard University Press, 1982: 84–106.
- Shearmur, Jeremy. *Hayek and After: Hayekian Liberalism as a Research Program*. New York: Routledge, 1996.
- Shklar, Judith N. *Men and Citizens: A Study of Rousseau's Social Theory*. Cambridge: Cambridge University Press, 1969.
- Shweder, Richard, Manamohan Mahapatra, and Joan G. Miller. "Culture and Moral Development." In *The Emergence of Morality in Young Children*, edited by Jerome Kagan and Sharon Lamb. Chicago: University of Chicago Press, 1987: 1–83.
- Sidgwick, Henry. *The Methods of Ethics*, 7th ed. Chicago: University of Chicago Press, 1962.
- Simon, Herbert. *The Sciences of the Artificial*, 3rd ed. Cambridge, MA: MIT Press, 1996.
- Skinner, Andrew. "Natural History in the Age of Adam Smith." *Political Studies*, vol. 15 (1967): 32–48.
- Skinner, B. F. *Walden Two*. Indianapolis: Hackett, 2005.
- Skitka, Linda J., Christopher W. Bauman, and Edward G. Sargis. "Moral Conviction: Another Contributor to Attitude Strength or Something More?" *Journal of Personality and Social Psychology*, vol. 88 (2005): 895–917.
- Skyrms, Brian. *The Evolution of the Social Contract*. Cambridge: Cambridge University Press, 1996.
- Skyrms, Brian. *Signals: Evolution, Learning and Information*. New York: Oxford University Press, 2010.
- Skyrms, Brian. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press, 2004.
- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*, edited by Edwin Cannan. Chicago: University of Chicago Press, 1976.
- Smith, Adam. *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L. Macfie. Indianapolis: Liberty Fund, 1982.
- Smith, Peter. *Explaining Chaos*. Cambridge: Cambridge University Press, 1998.
- Sober, Elliot, and David Sloan Wilson. *Unto Others: The Evolution and Psychology and Unselfish Behavior*. Cambridge, MA: Harvard University Press, 1998.

- Solow, Robert M. "Economic History and Economics." *The American Economic Review*, vol. 75 (1985): 328–31.
- Soros, George. "Fallibility, Reflexivity, and the Human Uncertainty Principle." *Journal of Economic Methodology*, vol. (2013): 309–29.
- Stanford, Craig B. *The Hunting Apes: Meat Eating and the Origins of Human Behavior*. Princeton, NJ: Princeton University Press, 1999.
- Stanford, Kyle. "The Difference between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation." *Behavioral and Brain Sciences*, vol. 41 (2018): 1–13.
- Sterelny, Kim. "Cooperation, Culture, and Conflict." *British Journal of the Philosophy of Science*, vol. 67 (2016): 31–58.
- Sterelny, Kim. "The Return of the Group." *Philosophy of Science*, vol. 63 (December 1996): 562–84.
- Stiner, Mary C. "Carnivory, Coevolution, and the Geographic Spread of the Genus *Homo*." *Journal of Archaeological Research*, vol. 10, no. 1 (March 2002): 1–63.
- Stiner, Mary C., Ran Barkai, Avi Gopher, and James F. O'Connell, "Cooperative Hunting and Meat Sharing 400–200 KYA at Qesem Cave, Israel." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 32 (August 11, 2009): 13207–12.
- Stuntz, William J. "Self-Defeating Crimes." *Virginia Law Review*, vol. 86 (2000): 1871–99.
- Sugden, Robert. *The Economics of Rights, Co-operation and Welfare*. Oxford: Blackwell, 1986.
- Sunstein, Cass R. *How Change Happens*. Cambridge, MA: MIT Press, 2019.
- Suwa, G., B. Asfaw, R. T. Kono, D. Kubo, C. O. Lovejoy, and T. D. White. 2009. "The *Ardipithecus ramidus* Skull and Its Implications for Hominid Origins." *Science*, vol. 326 (2009): 68–68e7.
- Tajfel, Henri, M. G. Billig, R. P. Bundy, and Calude Flament. "Social Categorization and Intergroup Behavior." *European Journal of Social Psychology*, vol. 1 (1971): 149–78.
- Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007.
- Taleb, Nassim Nicholas. *Foiled by Randomness: The Hidden Role of Chance in Life and in the Markets*. New York: Random House, 2004.
- Talisse, Robert B. *A Pragmatist Theory of Democracy*. New York: Routledge, 2007.
- Tenner, Edward. *Why Things Bite Back: Predicting the Problems of Progress*. London: Fourth Estate, 1996.
- Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: University of Princeton Press, 2005.
- Tetlock, Philip E., and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. New York: Crown, 2015.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*, rev. and expanded ed. New York: Penguin Books, 2009.
- Tiger, Lionel, and Robin Fox 1971. *The Imperial Animal*. London: Secker and Warburg.
- Tognato, Carlo, ed. *Cultural Agents Reloaded: The Legacy of Antanas Mockus*. Cambridge, MA: The Cultural Agents Initiative at Harvard University Press, 2017.
- Tomasello, Michael. *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press, 2016.

- Tomasi, John. *Free Market Fairness*. Princeton, NJ: Princeton University Press, 2012.
- Tooby, John, and Leda Cosmides. "Groups in Mind: The Coalitional Roots of War and Morality." In *Human Morality and Sociality: Evolutionary and Comparative Perspectives*, edited by H. Høgh-Olesen. New York: Palgrave Macmillan, 2010: 191–234.
- Trivers, Robert L. "Evolution of Reciprocal Altruism." *The Quarterly Review of Biology*, vol. 46 (March 1971): 35–57.
- Turchin, Peter. *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth*. Chaplin, CT: Beresta Books, 2016.
- Turiel, Elliot, Melainie Killen, and Charles C. Helwig. "Morality: Its Structure, Functions and Vagaries." In *The Emergence of Morality in Young Children*, edited by Jerome Kagan and Sharon Lamb. Chicago: Chicago University Press, 1987: 155–243.
- Turnbull, Colin M. *The Forest People*. New York: Simon and Schuster, 1963.
- Turner, Piers Norris, and Gerald Gaus, eds. *Public Reason in Political Philosophy*. New York: Routledge, 2018.
- Tyler, Tom R. *Why People Obey the Law*. New Haven, CT: Yale University Press, 1990.
- Vallier, Kevin. *Must Politics Be War? Restoring Our Trust in the Open Society*. Oxford: Oxford University Press, 2019.
- Vallier, Kevin, and John Thrasher. "The Fragility of Consensus: Public Reason, Diversity and Stability." *European Journal of Philosophy*, vol. 23 (2015): 933–54.
- Van Schoelandt, Chad. *The Community of Public Reason*. Manuscript.
- Van Schoelandt, Chad. "Functionalist Justice and Coordination," *Social Theory and Practice*, vol. 46 (2020).
- Vanberg, Viktor. *Rules and Choice in Economics*. London: Routledge, 1994.
- Vanderschraaf, Peter. *Strategic Justice: Convention and Problems of Balancing Divergent Interests*. New York: Oxford University Press, 2019.
- Vaughn, Karen. 1999. "Hayek's Theory of the Market Order as an Instance of the Theory of Complex, Adaptive Systems." *Journal de Economistes et des Etudes Humaines*, vol. 9 (Juin–Septembre 1999): 241–56.
- Vedder, Richard. "The Flight to Quality in College Admissions." *Forbes*, March 9, 2020.
- Vehrencamp, Sandra L. "A Model for the Evolution of Despotic Versus Egalitarian Societies." *Animal Behavior*, vol. 31 (1983): 667–82.
- Voors, Maarten J., Eleonora E. M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review*, vol. 102 (2012): 941–64.
- Wagner, Christian, Sesia Zhao, Christopher Schneider, and Huaping Chen. "The Wisdom of Reluctant Crowds," HICSS, 43rd Hawaii International Conference on System Sciences, 2010: 1–10.
- Waldron, Jeremy. *The Dignity of Legislation*. Cambridge: Cambridge University Press 1999.
- Waldron, Jeremy. *Law and Disagreement*. Oxford: Oxford University Press, 1999.
- Waldron, Jeremy. "A Right to Do Wrong." *Ethics*, vol. 92 (1981): 21–3.
- Waldrop, M. Mitchell. *Complexity: The Emerging Science as the Edge of Order and Chaos*. New York: Simon and Shuster, 1992.
- Weber, Max. *General Economic History*, translated by Frank Knight. New York: Collier, 1961.
- Weisberg, Michael, and Ryan Muldoon. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science*, vol. 76 (2009): 225–52.

- Westermarck, Edward. *Ethical Relativity*. Westport, CT: Greenwood Press, 1970.
- Weymark, John. "Cognitive Diversity, Binary Decisions, and Epistemic Democracy," *Episteme*, vol. 12 (2015): 497–511.
- Whitman, Douglas Glen. "Hayek contra Pangloss on Evolutionary Systems." *Constitutional Political Economy*, vol. 9 (1988): 450–66.
- Wiens, David. "Demands of Justice, Feasible Alternatives, and the Need for Causal Analysis." *Ethical Theory and Moral Practice*, vol. 16 (2013): 325–38.
- Wiens, David. "The Tyranny of a Metaphor." *Cosmos + Taxis*, vol. 5 (January 2018): 13–28.
- Wiessner, Polly. "From Spears to M-16s: Testing the Imbalance of Power Hypothesis among the Enga." *Journal of Anthropological Research*, vol. 62 (Summer 2006): 165–91.
- Williams, George C. *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press, 1966.
- Williamson, Oliver E. *The Economic Institutions of Capitalism*. New York: Free Press, 1985.
- Wilson, David Sloan. *Does Altruism Exist? Culture, Genes and the Welfare of Others*. New Haven, CT: Yale University Press, 2015.
- Wilson, David S. "Two Meanings of Complex Adaptive Systems." In *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, 2017: 31–46.
- Wilson, Edward O. *On Human Nature*. Cambridge, MA: Harvard University Press, 1978.
- Wilson, Edward O. *The Social Conquest of the Earth*. New York: Liveright, 2012.
- Witt, Ulrich. 1994. "The Theory of Societal Evolution: Hayek's Unfinished Legacy." In *Hayek, Co-ordination and Evolution*, edited by Jack Birner and Rudy van Zip. London: Routledge: 178–89.
- Wolpert, David H., and William G. Macready. "No Free Lunch Theorems for Optimization." *IEEE Transactions on Evolutionary Computation*, vol. 1 (April 1997): 67–82.
- Wolpert, David H., and William G. Macready. "No Free Lunch Theorems for Search." Report SFI-TR-9502-010. Santa Fe Institute, 1996.
- Woodend, Ashleigh, Vera Schölmerich, and Semiha Denктаş. "'Nudges' to Prevent Behavioral Risk Factors Associated with Major Depressive Disorder." *American Journal of Public Health*, vol. 105 (November 2015): 2318–21.
- WorldBank. "Decline of Global Extreme Poverty Continues but Has Slowed: World Bank." 2019. <https://www.worldbank.org/en/news/press-release/2018/09/19/decline-of-global-extreme-poverty-continues-but-has-slowed-world-bank>.
- World Bank. World Development Report 2012: *Gender Equality and Development*. Washington, DC: World Bank Publications, 2012.
- World Bank. World Development Report 2015: *Mind, Society, and Behavior*. Washington, DC: World Bank Publications, 2015.
- World Bank. World Development Report 2017: *Governance and the Law*. Washington, DC: World Bank Publications, 2017.
- Wrangham, Richard W. "An Ecological Model of Female-Bonded Primate Groups." *Behaviour*, vol. 75 (1980): 262–300.
- Wrangham, Richard, and Dale Peterson. *Demonic Males: Apes and the Origins of Human Violence*. Boston: Houghton Mifflin.
- Xiao, Erte. "Justification and Conformity." *Journal of Economic Behavior & Organization*. vol. 136 (2017): 15–28.
- Young, Iris Marion. *Responsibility for Justice*. Oxford: Oxford University Press, 2011.

Index

- abortion, 105, 162, 175–6
Abraham, Anita, 132
accountability, 52, 54, 58–9, 129, 135, 154,
159, 161–2, 236–7
Ache of Paraguay, 42–3, 132, 138
adjacent possible, 15, 113, 115–19, 127, 130,
136, 142–3, 194, 197, 245
agent-based models, 228–30
agriculture, 31, 75, 78
Alexander, Richard D., 52
Alf, *passim*
alphas, 23–7, 81
altruism, 37, 72, 76, 85, 170
anarchy, 146, 191
ancestral *Pan*, 21–34, 37, 40, 47, 50, 76–80,
84, 88–9, 91
Appiah, Kwame Anthony, 159
archaic state, 78–80, 138
Arthur, W. Brian, 114, 126, 128, 150
assurance problems, 179, 181
authoritarianism, 3, 118
autocatalysis, 14, 64, 102, 114, 118, 129, 134,
136, 165, 170, 195, 207, 238, 247–8
autonomy, 6, 14, 24, 36–7, 50–1, 54, 63, 81,
133, 138–9, 166, 201, 234
Axelrod, Robert, 38, 67, 196

Baier, Kurt, 88–9, 163
Bannerjee, Abhijit V., 216, 232, 239–40
Bahrami-Rad, Duman, 82–3
Beauchamp, Jonathan, 82–3
behavioral phylogenetics, 21, 26
Beinhocker, Eric D., 110, 112, 119, 185
Bellah, Robert, 80
Betty, *passim*
Bicchieri, Cristina, 49–50, 69, 86, 234–6
Binmore, Ken, 15, 38
Black Swan event, 188
Bloomington School, 203, 228
Boehm, Christopher, 22, 24, 31, 33, 35, 42, 48,
52, 77, 80
Bogota, 173–4
bonobos, 21–2, 25

Bowles, Samuel, 46
Boyd, Robert, 41, 69, 74
Bridgeman, Guy A., 189
Buchanan, James, 146, 248
Burke, Edmund, 197
butterfly effect, 188, 208

Cannan, Edwin, 110
Carpenter, Malinda, 76
Cartwright, Nancy, 231
cascades, 57, 127, 237
categories, 105–9, 115, 117–18, 242–3
Catholic Church, 82–3
causal opacity, 88, 95, 101, 164, 227
central planning, 7, 13n7
Cephu, 44–6, 51–5, 77, 164
chaos, 120–2, 126–7, 188, 223n42
chimpanzees, 21–9, 66
ciguatera poisoning, 65–6
cities, 204–5
Civil Rights Act of 1964, 198
clans, 36, 75, 82, 90
closed society, 3, 35, 69, 76, 81, 86–7, 102, 207
closure rules, 135–6
coalitions, 23, 54
coevolution, 30, 64–5, 70–1, 76, 116,
126, 128
cognitive biases, 74, 106, 211, 217
Cohen, Alma, 190, 196
Colander, David, 233, 241
collective action problems, 202
combination, evolution by, 114
Common Ancestor, 21–4, 27
common good, 178–9
common pool resources, 187, 199, 202, 227–8
communitarianism, 91, 104n5, 205
Community-Driven Development, 233
Complex Society Moral Package, 138,
159–62, 195
complex system, 7–8, 10–13, 15, 119, 122–3,
127, 142, 164, 184, 187–8, 194–5, 200,
203, 205, 208–9, 211, 213, 218, 220–2,
228, 231, 238, 240–4

- complexity, 6–15, 88, 102–3, 118–24, 126–30, 164, 171–2, 185–7, 192–5, 197–8, 200, 204–5, 208–9, 212–14, 216, 218, 221, 223–8, 230, 233–4, 237–8, 241–5, 248
 complexity catastrophe, 122, 126
 compliance, 35, 47, 49, 52, 59, 69, 90, 173, 177–9
 complicated determinate system, 210–11, 213
 conditional cooperators, 40–3, 48, 53–4, 70, 138, 149, 153, 181
 conservatism, 9–10, 82, 102, 105, 136, 165, 197, 223, 234
 constructivism, 7, 199
 conventions, 88, 160–2, 181
 convergence, 22, 25n35, 56–7, 60, 63, 145–6, 146n18, 150–60
 cooperation, 14, 25, 37–8, 40, 42–3, 47, 49, 57, 59, 72, 80–1, 90, 102, 129–30, 161, 199, 247
 coordination, 27, 49, 52, 54–5, 63, 111–12, 148, 150, 161, 202
 coordination problems, 179, 181
 coordination-reconciliation game, 55, 63, 68, 70, 145–7
 copying, 66–8, 102, 113, 115, 118
 Cosmides, Leda, 55, 60, 62, 150
 cosmopolitan, 81–2, 91
 creative destruction, *gales of*, 121, 123, 125, 128, 195–6, 248
 creativity, 66
 Cuban Missile Crisis, 215
 cultural evolution, 6, 8, 66, 68–70, 74–5, 81, 101
 cultural relativism, 62, 68
 Cummins, Denise, 26

 dancing landscapes, 224–5, 239
 Darwin, Charles, 20, 47, 115
 de facto vs de jure morality, 88–9
 De Waal, Frans, 23
 decomposability, 122–23, 127, 187–8, 204–5, 207–8
 Deliberative Model, 144–7, 150, 154, 157
 democracy, 3, 5, 12–13, 15, 51, 99, 132, 137, 172–3, 178, 221, 241, 245–6
 demoralization, 62, 146
 Denktas, Semiha, 217
 deontic reasoning, 26
 DeScioli, Peter, 53–4, 59, 61, 150
 despotism, 23, 26, 29–30, 91, 138
 development, 109, 191–4, 232, 240
 Dewey, John, 95–6, 99–100, 106–7, 186, 221, 227
 diminishing returns, 120
 diversity, 3, 7, 9, 13–15, 20, 60, 68–9, 97–9, 102–5, 109, 111–14, 116, 118–21, 129, 134, 136, 140–1, 143, 146, 148, 151, 153, 156–62, 164, 167, 170, 195–7, 204–7, 213, 217, 227, 238, 240, 243–4, 247–8
 diversity theorems, 201, 213
 division of labor, 30, 41–2, 97, 110–12, 118, 134, 170
 domination, 23, 26, 32–4, 36–7, 48, 84, 91, 166
 Duflo, Esther, 216, 232, 239–40
 Durbin, J., 189–90
 Durkheim, Emile, 210

 egalitarianism, 4, 14, 17, 19, 22, 24, 30–7, 43, 46, 48, 76–80, 84–6, 88, 91, 131, 133, 247
 Einav, Liran, 190
 Einstein, Albert, 115
 emergence, 9n17, 124, 171, 192–3
 enforcement, primary and secondary, 190
 epistocracy, 244–5
 equality, 3, 22, 29, 32–3, 35–6, 51, 78, 80–1, 96, 100, 166
 equilibrium models, 11, 120, 181
 equilibrium-stabilizing mechanism, 57
 error inflation, 208–9, 214
 Estlund, David, 158
 ethnic markers, 70, 76, 80, 84, 90, 112
 European Union, 174, 226
 evolutionary explanation, 5, 11, 20
 exchange, 41, 109–10, 112, 130
 exogenous influences, 12, 127, 158, 178, 189, 199, 204–5, 220, 225, 227, 233, 246
 experimentalist governance, 226, 230, 240
 experiments in living, 97–8, 106, 230
 experiments, social, 173, 186, 193
 experts, 13, 13n8, 15, 101n32, 121, 185, 188, 191, 211–13, 216, 241–5
 exploration-exploitation trade-off, 66–7
 externalities, 134, 180n28, 203–4, 245, 256
 Exxon Valdez, 214

 fairness, 33, 54, 82, 83
 fallibility, 96, 138
 Farmer, Charles M., 190
 Fehr, Ernst, 85–6

- Ferguson, Adam, 19
 Flannery, Kent, 77–8
 flexible conscience, 48–9, 54, 63, 73, 84,
 134n25, 138
 foxes, 212–13
 fragility and anti-fragility, 196–7
 Free Lunch Theorems, 243
 free-riders, 46, 49, 125, 181
 Freeden, Michael, 10n20, 96

 Garbacz, Christopher, 189–90
 garden city movement, 205
 Gauthier, David, 163, 166
 Geertz, Clifford, 106
 gender, 3, 36, 81, 100, 102
 gene-culture evolution, 8, 30, 64–6, 70–1, 76
 general will, 178–9
 Gilligany, Michael J., 77
 Gintis, Herbert, 46
 governance, 12, 14, 182, 194, 204, 207, 216,
 241, 248
 Governance Process, 178
 great society, 3
 Greene, Joshua, 162–3
 Gressman, Philip T., 229–30
 group selection, 4–6, 30, 70–2, 75, 124,
 166, 193
 group size, 77–8
 Gurven, Michael, 42

 Haidt, Jonathan, 53, 61, 81–2
 Hamilton, W.D., 30n2
 Hardin, Russell, 44
 Hardy, Jeremy, 231
 Hare, Brian, 28
 harm, 35, 47, 82–3, 104, 130n7, 136–7, 195
 Harvey, A.C., 189–90
 Hayek, F.A.
 Cartesian Rationalism, 88, 95, 99, 242, 244
 categorization, 108–9
 central planning, 7, 13n7
 closed society, 35
 complex systems, 7, 9, 11, 123, 184, 194,
 197, 204, 240
 complexity, x, 14
 constructivism, 7, 19, 140, 199, 241
 cultural evolution, 7–8, 65, 73, 95
 deontology, 166
 egalitarianism, 4, 14, 15, 84–5, 91
 evolution, 4–6, 11
 extended order, 5
 gene-culture evolution, 8, 64
 governance, 15, 204
 Great/Open Society, 3, 5–7, 10, 12, 14, 20–1,
 35, 37, 83–4, 86, 91, 95, 102, 124, 129,
 136, 138–9, 166, 184, 197, 240, 246–8
 group selection, 4–6, 70, 72–3, 122–4, 164
 holism, 72
 Human as rule-following animal, 9, 64, 129
 human nature, 96
 limits of reason, 9–10, 12, 14, 241
 local knowledge, 236
 moral intuitions, 12–13, 85
 nature-culture, 64, 66
 opacity of rules, 95, 102, 239
 order of actions, 72, 228
 prediction, 11, 194
 purpose-independent rules, 8, 87–8,
 136, 163–4
 rule-following, 9, 166, 194
 rules, abstract, 4
 rules, flexibility of, 73
 scientism, 11n4, 88, 244
 Scottish Enlightenment, 19, 20, 95
 small groups, 16, 21, 87, 132–3, 138
 social contract, 7–8, 140
 social learning, 66
 spontaneous order, 6, 228, 246
 synoptic delusion, 7, 242
 tradition, 167
 tribalism, 4–5, 11, 14, 132–3, 138
 utilitarianism, 165
 Heine, Steven J., 81, 106
 Herrmann, Esther, 28
 heterogeneity, 120
 Hi-Lo bias, 154–5, 157
 hierarchy, 19, 21, 23–6, 32, 47, 75, 78–80,
 84–5, 88, 112, 245
 hinder hindrances to self-organization,
 196, 237
 Hobbes, Thomas, 19–20, 30n1, 48, 129–30, 198
 Holocene Era, 24, 31, 75
 homologies, 22n10
 homosexuality, 105
 Hong, Lu, 201, 243–4
 Houston, David J., 189–90
 humans, rule-following animals, 9
 Humboldt, Wilhelm von, 96
 Hume, David, 130, 148, 156, 181
 Hungary, 174

- hunter-gatherer societies, 17, 20–1, 31, 33,
36–7, 40, 42, 46, 48, 50, 75–7, 79, 81,
109–10, 138
- imitation, 9, 66, 76, 90, 101, 118, 170
- impartiality, 54–5, 58, 60, 70, 80–3, 85, 90,
103, 133, 137, 140–1, 143–4, 146, 149–50,
160, 245, 247–8
- impatient moralist, 234–6
- incentives, 179, 181–2, 216, 222, 236
- incest, 82
- inclusivity, 3–4, 103, 133, 248
- increasing returns, 120, 150
- incrementalism, 188
- individualism, 35, 163–4, 170, 195
- inequality, 19, 22, 29, 75, 77–8, 80–1, 102,
110, 139
- innovation, 66–8, 112–19, 121, 193, 214,
225, 248
- institutional transplantation, 192–3, 200
- invisible hand, 39, 124
- Ismael, Jeann, 170, 172, 179, 184
- Jacobs, Jane, 204–5
- Jenness, Desmond, 36
- Jevons, W.S., 115
- Johnson, Neil, 186, 221, 227
- justification as advocacy, 59
- Kant, Immanuel, 53, 147, 149, 159
- Kauffman, Stuart, 15, 113, 115–17, 121–2,
126–8, 170
- Keynes, John Maynard, 191
- kin selection, 76
- kin altruism, 30
- Kitcher, Philip, 96
- Knight, Jack, 186, 221, 227
- Kuhn, Thomas S., 98, 108
- Kukathas, Chandran, 206
- Kupers, Roland, 233, 241
- Kurzarbeit*, 195
- Kurzban, Robert, 54, 59, 61
- Laland, Kevin N., 67
- landscapes, epistemic, 98–9, 222–4
- Lasswell, Harold, 23
- Late Pleistocene Era, 21, 23–4, 30–2, 35–7,
40, 42–4, 51, 63–4, 72, 75, 79–81, 84–6,
91, 109–10, 123, 247
- Lee, Richard Borshay, 34–5
- Legal Centralism, 172–3
- Legislative Theory of Democratic Self-
Governance, 172, 182
- levers, 11–12, 185, 214, 216, 218
- Lewis, Paul, 9n17
- liberal archipelago, 206–7
- liberal democracy, 3
- liberal pluralism, 207
- liberalism, Millian, 14, 95, 98–100, 103
- liberty, 82–3, 95, 113, 134, 137, 182, 195
- Lincoln, Abraham, 48
- Lindblom, Charles, 221–3, 227
- lock-in, 37, 129, 224, 238
- Lopez-Perez, Raul, 86
- Lorenz, Konrad, 5n10
- loyalty signalling, 62
- Lund, Adrian K., 190
- Machiavellian reasoning, 25, 30, 37, 39–40,
43, 48, 53, 79–80, 84, 125, 170
- Mackie, Gerry, 173
- macro, meso, micro, 183, 238
- macro selection, 73, 122–3, 126, 158, 164,
187, 197
- Malthus, Thomas, 20n5
- manipulation, 184–8, 194, 214–18, 220–1, 241
- Mapuche, 65
- Marcus, Joyce, 77–8
- market, 9n19, 31, 39, 41, 85–6, 109–11, 113,
115, 118, 126, 131–2, 197
- Martin, Christopher Flynn, 25n35
- Marx, Karl, 210
- McCloskey, Deirdre Nansen, 113
- Melis, Alicia, 28
- Members of the Public, 144–5
- Mercier, Hugo, 50, 52–3, 59
- Mesopotamia, 78
- micro governance, 184
- Mill, John Stuart, 97, 101n31, 106–7, 117, 137,
162, 167, 192, 230–1
- Miller, John H., 187
- mind-reading, 25, 52
- minimal state, 206
- Mockus, Antanas, 173–4, 235
- Modern Egalitarian Moral Package, 17, 54,
63, 68, 80, 86, 138–9, 163
- modularity, 186, 193, 204, 238
- moral intuitions, 12–13
- moral order, 20, 48
- moral progress, 96–100, 102–3, 106–7, 146

- moral relativism, 160, 162, 206
 moral theory (vs moral philosophy), 89, 96
 moral warfare, 62
 muddling. *See* tinkering
 Muldoon, Ryan, 97
 Müller-Lyer illusion, 217
 multilevel selection, 30, 71–2, 74, 78
 multiple regression, 208–9, 213, 222
 Mussolini, Benito, 245
 myside bias, 53, 62
- Naïve Naturalism, 107–8
 Nash equilibrium, 112, 145
 Natural Liberty, 134–7
 natural selection, 20
 Neely, Grant W., 189
 negative feedback, 120
 network effects, 120
 New Yorkers, 110, 112
 Newton, Isaac, 114
 niches, 112, 116, 127–8, 165, 170, 228, 245, 248
 Nichols, Shaun, 134
 NK complexity, 121, 126, 187–8, 223
 nonlinearity, 194, 209, 213–14, 219–20, 238
 Norenzayan, Ara, 81, 106
 North, Douglas, 109, 130, 182, 186
 Nozick, Robert, 133, 142, 206
 nudges, 217
 Nunnar-Winkler, Gertrude, 47–8
- O'Connor, Cailin, 112
 objectivity of morality, 56–9, 62, 160–1, 163
 objectivity-induced convergence, 56–7
 offsetting behavior, 189–90
 Open Society, 5–7, 35, 83, 86, 90–1, 98–9, 101–3, 110, 112, 117–18, 123, 126–28, 130–1, 133–4, 136, 139–40, 142–4, 146–7, 157, 160, 162, 164–7, 170, 172, 188, 197, 203–4, 206–7, 238, 240, 244–6, 248
 optimal eligible set, 144–5
 optimization, 14, 21, 67–8, 85, 91, 123, 158–9, 187–8, 222–4, 247
 optimizing stance, 158–9
 optionality, 195–6
 Ostrom, Elinor, 15, 180, 199–202, 206, 227–8
 Ostrom, Vincent, 15, 199–200, 202, 206, 227, 241
 over-imitation, 66, 76
 Over, Harriet, 76
- Page, Scott, 201, 213, 242–4
 Pareto rule, 144, 198
 parsimony, principle of, 21
 Pasquale, Benjamin J., 77
 path-dependence, 86, 102, 119–20, 170, 209, 228
 Peck, Jennifer R., 229–30
 Peltzman, Sam, 189
 perspectives, 97, 103–9, 115, 117–19, 140–1, 145–6, 148, 159–60, 165, 167, 242–3
 Plato, 112, 143, 184
 Platteau, Jean-Philippe, 81, 132, 138, 195
 polarization, 61, 152–4, 159, 240
 policy state, 178
 polyarchy, directly deliberative, 226
 polycentric governance, 200–3, 205–6, 227–8, 239, 244
 Popper, Karl, 3, 11–12, 35, 86, 95, 97, 99, 112, 117, 248–9
 Popular Control, 172–3, 178
 positive feedback, 120
 poverty, 232–3
 Powell, Russell, 146, 248
 predictability, 10–12, 14, 24, 87, 104, 117, 129–30, 136, 142, 157, 165, 171, 182, 188–90, 193–4, 197, 208–14, 216–18, 220–1, 228, 230–1, 238, 240–3
 pressing problem-solving context, 199–200, 239
 prestige, 77–8
 Prisoner's Dilemma, 38, 41, 67, 125, 129, 181
 procedural justice, 141–2
 Process Theory of Democratic Self-Governance, 178
 prohibitions, 26, 53, 65–6, 134–8, 165, 177, 182, 198, 202, 235–7
 property, 29, 36, 130n7, 130–4, 137, 182, 191, 193, 195
 public goods, 77, 179–81, 199
 public justification, 14–15, 50–1, 57–60, 63, 74–5, 103, 162, 167, 170, 234–5, 246–7
 punishment, 43, 46, 54, 138
- randomized control trial, 231–3
 rational choice theory, 203
 rationalism, 14
 Rawls, John, 15, 48, 50, 89, 140–1, 157, 163, 166, 191, 240
 reactive emotions theory, 86
 rearranging effects, 203
 reciprocal altruism, 38

- reciprocity, 14, 17, 30, 37–40, 41n18, 42–3, 53, 71, 80, 109–10, 118, 125, 132–3, 138, 170
 reconciliation, 14, 48–9, 54–5, 58, 60, 63, 84, 138, 149–51, 153, 157–8
 reference group, 153–4, 158
 reflexivity, 119–20, 125, 129, 142, 150, 157–8, 165, 170, 181, 186, 189–91, 194, 198–200, 207, 209, 216–17, 221, 223, 225, 227–8, 238–9, 244–5
 religion, 100
 resilience, 195–7
 revenge effect, 214, 218, 221
 reverse domination, 34–7, 54, 77, 81, 132–3, 138, 245
 Rhoads, Steven, 189
 Richardson, Lilliard, 189–90
 Richerson, Peter J., 41, 69, 74
 Ridley, Matt, 114
 risk-taking, 197
 robust justification, 51, 95, 162, 167, 234, 236
 Rosenberg, Alex, 224
 Rostow, W. W., 191
 Rousseau, Jean Jacques, 19–24, 26, 29, 77–8, 84, 96, 178
 rugged landscapes, 222, 224, 227
 rule violators, 52
 rule of law, 174, 191–2
 rules
 abstract, 4, 7, 87, 160
 flexibility, 73, 84
 following, 14, 26, 64
 internalization, 47–8, 64
 opacity, 95, 102, 164
 purpose-independent, 8, 86–8
 self-defeating, 174–5, 177
 sensitivity, 49–52, 54, 57, 59–60, 63, 69, 74, 103, 138, 167
 Ryan, David L., 189

 Saari, Donald, 119–20
 Sabel, Charles F., 187, 226–7, 230
 Samii, Cyrus D., 77
 Schmidt, Klaus M., 85–6
 Schölmerich, Vera, 217
 Schulz, Jonathan F., 82–3
 Schumpeter, Joseph A., 13n7, 121, 195, 241, 248
 scientific progress, 95, 97–9, 101n32, 102, 107–8, 117
 Scott, James C., 79
 search heuristics, 227
 seat belts, 189–90, 216
 sectoral governance, 184, 198–9
 secularity, 100, 162
 self-control, 47
 self-governance, 170–1, 178, 183–4, 186, 202, 237
 self-interest, 37, 52, 84, 132, 170, 247
 self-organization, 7, 14–15, 124, 127–8, 147, 158–9, 164–5, 167, 170, 172–4, 177, 181–2, 187, 194–5, 197–8, 201–4, 228, 232, 237, 240, 245–7
 self-organized criticality, 127
 self-segregation, 62–3
 side-taking, 54, 58, 85, 90, 133
 Simon, Herbert, 122, 126, 187
 Skinner, B.F., 20n1
 Skyrms, Brian, 41
 small-scale bands, 4
 Smith, Adam, 41–2, 85, 109–10, 112, 118, 126, 134–6
 social contract, 7–8, 14, 19, 50, 140, 143, 157, 170
 social Darwinism, 5
 social learning, 29, 67, 101, 193
 social norms, 49
 social reform, 9, 15, 60–1, 89–90, 96, 184, 186–7, 239–40, 248
 Sodian, Beate, 47–8
 specialization, 111–12, 116, 170
 Sperber, Daniel, 52–3, 59
 spontaneous order, 6, 44, 143
 stable justification, 51, 58, 61, 145
 Stag Hunts, 41–2, 125, 129, 151, 181
 Stanford, Kyle, 55, 60–1
 Stiner, Mary, 32
 strategic dilemmas, 179, 181–3, 198–9, 202–3, 206, 227–8, 239, 244–5, 248
 strategic reasoning, 8, 25–26, 28, 48
 Strawson, P.F., 163
 strong reciprocators, 40–1, 43, 72, 125
 Stuntz, William J., 174
 suboptimality traps, 179
 subordinate resistance, 33, 113
 subsystem governance, 184
 superforecasters, 213–14
 superstition, 101, 107
 synoptic delusion, 7
 systemic governance, 184

- Taleb, Nassim Nicholas, 188, 195, 197
- Tenner, Edward, 203, 214, 218
- territoriality, 27
- testing conception, 89
- Tetlock, Philip, 211–14, 221
- tight coupling, 203
- tinkering, 114, 122, 221–7, 238
- tipping points, 194, 219, 230
- tolerance, 80, 97, 161, 164
- Tomasello, Michael, 25, 28, 59, 76
- Tooby, John, 55, 60, 62, 150
- Tostan Community Empowerment Program, 51, 236
- totalitarianism, 99
- trade, 109, 111
- tragedy of the commons, 236
- tribalism, 3–4, 6, 10, 14, 20, 35, 76, 81, 84, 90–1, 102, 131–3, 138, 162–4, 247
- Tribalistic Package, 90–1
- Trivers, Robert, 38–9
- trust, 42, 69, 76–7, 90
- Turchin, Peter, 77, 80
- Turnbull, Colin, 44–5
- Tyler, Tom R., 177
- Ultimatum Game, 85–6, 132
- ultra-sociality, 21, 30, 39–40, 48, 53, 85, 88–9, 91, 101–2, 109–10, 129, 170
- unintended consequences, 200
- utilitarianism, 165, 167
- utopia, 97, 162, 184, 191, 205–6, 248
- veil of ignorance, 140
- Voors, Maarten J., 77
- Watson, James, 115
- wealth, 113
- Weber, Max, 132
- weighting function, 148–53, 158
- Western Educated Industrialized Rich Democracies (WEIRD), morality of, 81–3, 85, 90–1, 137, 162
- Wilson, David Sloan, 122–24, 126
- Wilson, E.O., 5, 76
- wisdom of the crowd, 243
- Wonder, Stevie, 101
- Woodend, Ashleigh, 217
- World Bank, 177–8, 192
- Wright, Sewall, 5
- xenophobia, 27, 69, 76
- Xiao, Erte, 52
- Yanomano, 110, 112
- Yasawa Island, 65, 86, 101
- Young, Iris Marion, 173
- Zeitlin, Jonathan, 187, 226–7, 230

